THE POWER IN ASSERTION

Discursive Agency, Norms, and the Unity of Thought

Der Fakultät für Sozialwissenschaften und Philosophie
der Universität Leipzig

genehmigte

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Philosophie

Dr. phil.,

vorgelegt

von M.A. Yiwen Zhan

geboren am 09. November 1985 in Gansu, China

Gutachter: Prof. Dr. Sebastian Rödl
Prof. Dr. Pirmin Stekeler-Weithofer

Tag der Verleihung: 3. Jul 2018

Bibliographische Beschreibung

Zhan, Yiwen

The Power in Assertion: Discursive Agency, Norms, and the Unity of Thought

Universität Leipzig, Dissertation

XII, 237 S.

Schlagworte: Pragmatics. Speech Acts. Frege, Gottlob. Assertion. Normativity. Discourse Analysis. Philosophy of Language.

Yiwen Zhan defends a new account of the pragmatics of assertion, according to which assertions are agents' performative speech acts of commitment to truth. He explains how such a pragmatic approach can be fitted into Fregean context and account for the force–content relation, non-assertoric contents, context-sensitivity, constitutive norms, belief and the dynamics in discourse, and other related problems.

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten: Weitere Personen waren an der geistigen Herstellung der Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und ist auch noch nicht veröffentlicht worden.

Leipzig, den 08. Januar 2018

Yiwen Zhan

To my mother, Li Li,
and to my father, Jibao Zhan

Preface and Acknowledgements

This dissertation is an essay on the pragmatics of assertion. It derives from a former project on normativity in Hegel's Logic of the Concept. This might sound like a drastic change in topic. But I think the underlying philosophical idea is not that different—Readers familiar with Hegelian philosophy may find out that many themes in this essay echoes Hegel's idea of Logic: For instance, my favored dynamic approach to constitutive agency introduced in Chapter 5 and 6 is reminiscent of Hegel's idea that the Logic is the self-movement of the Concept; and the criticism of Kantian constitutivism and the contention that our agency is social-historically mediated clearly reflects a Hegelian view towards the relation between the Logic and the Philosophies of Nature and Spirit.

I originally attempted to reconstruct Hegel's arguments in his Logic based on these themes. Yet later I found out that this would be an enormously difficult task to take on. It's not just because Hegel's language is famously idiosyncratic, but also, and more importantly, because I always want my interpretation of Hegel to be accessible not only by Hegel scholars but by audiences of philosophical discussions at large—This is not for advertising purposes; rather, I believe it's beneficial for myself to learn to conduct examinations of valuable ideas in Hegelian philosophy or any historical philosophical idea in general.

Philosophy has gained significant developments in topics concerning logic in the past century or so, especially under the influence of Frege's project of the Begriffsschrift. However, if I force to rehearse Hegel's own arguments in his doctrine of judgment and meanwhile also try to carry out investigations and assessments of the contemporary understanding of judgment and assertion, the essay would rather appear to be a weird and arbitrary juxtaposition of philosophical doctrines, which, I am afraid, both Hegel scholars and non-Hegelian readers would hardly find satisfying—After all,

I'm by no means a seasoned philosopher, and it would be too challenging for me to achieve anything close to this standard.

Therefore, I eventually gave up on trying to write a Hegel essay. Although there may well be certain Hegelian imprints in my point of view, this essay deals with the pragmatics of assertion in a general sense. It is nevertheless an ambitious and challenging project for me, and whether it has succeeded in providing any interesting account of assertion is left to the readers to decide.

My biggest inspiration in writing this essay comes from my two dissertation advisors: Sebastian Rödl and Pirmin Stekeler-Weithofer. I've been attracted by Sebastian's work since before I came to Leipzig to pursue my doctoral study. And during these years, his advice and criticism have been the most important intellectual stimulations to me. I would like to thank Sebastian for his crucial help in finishing this dissertation. I'm also greatly indebted to Pirmin for his suggestion and encouragement. I've benefited a lot from his insightful works—especially his readings of Hegel. Sebastian and Pirmin have taught me how to learn from the history of ideas and at the same time philosophize on my own. Their passions for philosophy have kept me motivated and humble. Without their influence, I don't think I could finish a dissertation as challenging as this one.

I've been fortunate to have also received lots of help from many excellent philosophers and friends over the years. Thanks to Günter Zöller and Christian Martin, who read the first drafts of my dissertation proposal and gave me very important advice. I owe them a great debt for their help. Thanks to Florian Ganzinger and Wei Cheng, who read parts of the earlier draft of my dissertation and gave me tons of precious feedbacks. Thanks also to Jonas Held, who read the outline of this essay and gave me many valuable suggestions for making the essay more readable. Furthermore, I'm especially grateful to Wolfram Gobsch and Matthias Haase. I've benefited greatly from the numerous conversations with them in and out of class, which have helped me gain confidence in philosophy.

I'm thankful for having had the opportunities to give talks on my earlier project at two workshops in Leipzig: A workshop on absolute idealism with Paul Franks and Sebastian Gardner in 2014; And a workshop on the form of practical knowledge with Stephen Engstrom in 2015. I received many inspiring comments crucial for my argument in the 5th chapter of this dissertation.

This dissertation received most feedbacks and objections from the colloquiums at the Institute of Philosophy, University of Leipzig, and earlier at the University of Munich, at which I have presented my project and multiple chapters. Thanks to all the participants of those colloquiums and to all my interlocutors on the related topics over the years; in particular, thanks to Javier Álvarez-Vázquez, Sebastian Böhm, Carolin Böse-Sprenger, Robert Brandom, James Conant, Stephen Engstrom, Michael Frey, Rebekka Gersbach, Adrian Haddock, Tobias Kasmann, Andrea Kern, Thomas Khurana, Irad Kimhi, Chencheng Liu, Gino Margani, Alexandra Newton, Lars Osterloh, Martin Palauneck, Nicole Del Rosario, Thomas Spiegel, and Jana Thesing. I'm sure I'm forgetting someone.

The China Scholarship Council (CSC) has financially supported my doctoral study. I would like to express my gratitude for the support.

My life over the past few years has been almost solely occupied by my work on this dissertation. I'm fortunate to be able to finish my work in such calm and uninterrupted period of time. The calmness, nonetheless, is filled with stress and anxiety—These years give me an increasingly uneasy feeling about the future. I'm not sure, though, how much of the unease is because of the changes in the world and how much is just because of the changes in myself; just like I'm not sure whether I should have questioned myself more on what to do or rather less—A career in academic philosophy, as it appears to me, offers inspiration from great minds and hopefully a standard of living. It is a means to achieve good. But it does not prevent you from feeling weak and ashamed, and it offers less comfort if you do feel so. Besides, a PhD life has the risk of distorting your prohairesis, taking most of the fun out of your research, and, perhaps, inducing mild brain damage. I'm lucky that many people, including many 'philosopher friends' mentioned earlier, have supported me through the tough times, which has reduced that risk greatly. Thanks also to many other friends who don't work in academic philosophy. Their support is equally important to me, and I believe they will forgive me for not mentioning their names. But in particular, I would like to express my special thanks and also apologies to Dadao Hou, Wibhu Kutanan, Jiande Li, and Yue Wu. You have kept listening and tolerated me unleashing 'negative energy' in the dimmest days. Thanks, and sorry for that.

I want all my family and friends to know that although I've had little time to spend with you while I was wrestling with my dissertation, it is your trust and confidence in me that keep me going along and trying to become a better person.

Most of all, thanks to my parents. The completion of this essay would be impossible without their support and understanding. This essay is dedicated to them.

Yiwen Zhan

Jan. 6th, 2018

* * *

“ [...]Aber das Irrsal
Hilft, wie Schlummer, und stark machet die Not und die Nacht,
Bis daß Helden genug in der ehernen Wiege gewachsen,
Herzen an Kraft, wie sonst, ähnlich den Himmlischen sind.
Donnernd kommen sie drauf. Indessen dünket mir öfters
Besser zu schlafen, wie so ohne Genossen zu sein,
So zu harren, und was zu tun indes und zu sagen,
Weiß ich nicht, und wozu Dichter in dürftiger Zeit.
Aber sie sind, sagst du, wie des Weingotts heilige Priester,
Welche von Lande zu Land zogen in heiliger Nacht.”

— Friedrich Hölderlin

Contents

Detailed Contents

Introduction

In this dissertation, I explore the significance of the assertoric force in thinking, i.e., in the linguistic-discursive practice of humans. Specifically, I explore the pragmatic significance behind the conception of force for assertion-making and explain how a pragmatic approach can be fitted into Fregean context and account for the force–content relation, non-assertoric contents, context-sensitivity, constitutive norms, belief and the dynamics in discourse, and other related problems.

According to the modern speech act theory, assertions, together with requests, promises, warnings, apologies, etc., indicate different modes of performative utterances and hence different forces. However, it is notable that when Gottlob Frege first conceived assertions to be indicating a type of force, he took the assertoric force to be *the only* performative force that can be written in his Concept Script (the *Begriffsschrift*)—written as the so-called 'judgment-stroke'—For him, the assertoric force is *the* force for assertion, which expresses a thought as thought, i.e., as intellectually universally sharable content.

This dissertation will not serve as an interpretation of Frege's own writings. It intends instead to reexamine the relation between thought and the power of thinking based on the Fregean theme originated from his conception of the assertoric force—Frege's conception of the singularity of the force, I think, roots deeply in the broadly rationalist tradition in philosophy (e.g. Kant) that there must be a singular power of Reason for logical thinking, namely the faculty of Reason (or similarly, a singular 'space of reasons') that enables us to think and communicate in a language governed by a universal logical form. However, for us contemporary readers to comprehend and to appreciate the traditional rationalist idea of the power of Reason through the deeply influential Fregean lens of truth-conditional semantics, the influence of Frege and the so-called 'analytic tradition' in the last century has brought us both advantages and disadvantages: The evident advantage of the influence is that it highlights the importance of logic and conceptual analysis in our understanding of basic philosophical issues; The disadvantage, however, is that Frege's logicist approach is deeply

mathematically oriented, which tends to favor a minimalist approach to the general semantics of conceptual contents, which undermines the significance of the performative force or the power of thinking—According to the minimalist approach, while the contents and their meaning enjoy the sui generis feature of universal intelligibility, our practice of thinking and discourse inherently belongs to our psychological life, which confers meaning only by chance.

On the contrary, I will argue that our pragmatic, assertoric performance is meaning-constitutive, hence the minimalist approach must be rejected. The assertoric force, therefore, should not be deemed as simply tantamount to the supposedly autonomous semantics; rather, the force reveals a crucial aspect of our conceptual thinking: the human agency and its pragmatic, discursive commitment. From this aspect, I will discuss in this dissertation how we can better understand the singularity of the force and the nature of content—the insightful observation that Frege originally made but failed to live up to in his project—and how the singularity of the force could be compatible with the plurality of life and meaning.

To develop the argument, I will indicate that the assertoric force plays a vital role in understanding two important philosophical questions, which I call the *question of demarcation* and the *question of distanced analyzability*, respectively:

The question of demarcation concerns the condition of intelligibility of a thought. It asks how it is possible to demarcate thoughts from other occurrences or constituents of our mental lives; or equally, how to demarcate an intellectually sharable content from other intrinsically unintelligible contents—A straightforward Fregean answer, then, is roughly that we can demarcate a thought if and only if it can be asserted by the assertoric force.

The question of distanced analyzability, in turn, asks how it is possible to distance ourselves from the assertion of a thought without giving up the condition of intelligibility of that thought. In other words, according to the project of *Begriffsschrift* the assertoric force always indicates that the truth (or the truth-condition) of the content is recognized in form of a judgment or an assertion; yet there appears to be many ways, as Frege himself admits, to think or communicate contents other than asserting it (e.g. when we present a thought in an indirect speech). This dissertation will show that these two questions are interrelated, and needs a unified answer.

It might seem that conflicting approaches are needed to account for these two questions. In his writings, we can see that Frege had also been struggling to reconcile

the tension between the *product* of logical thinking (judgments, i.e., essentially true thoughts) on the one hand and the *act* of thinking and judging on the other—a tension, so to speak, that seems to be residing in the characteristic finitude of human intellect, i.e., in the fact that we are neither logically omniscient nor simply stupid but have the discursive or epistemic power to understand things and make our knowledge grow. In my dissertation, I show that the question of demarcation and the question of distanced analyzability are interrelated and demand a unified answer. To resolve the tension that Frege had been struggling with, I also indicate that we need to reconceive the relation between pragmatics and truth-conditional semantics and the idea of the power of rational thinking.

Frege's Force–Content Distinction

In Chapter 1, 'Frege's Assertoric Force', I begin by putting forward the question of demarcation, followed by an introduction of Frege's basic conception of the assertoric force. I suggest that we can read Frege's invention of the sign for assertoric force in *Begriffsschrift* as an attempt to introduce the 'demarcator' of contents. An important consequence of this idea, I argue, is that by being incorporated into the *Begriffsschrift* as the demarcator of contents, the interpretation of force cannot be inflationary—The force thus cannot be assimilated to a kind of content: It neither makes any contribution to the content nor alters it; rather, as Frege consistently emphasized, by asserting a content (as essentially true) the force demarcates the content as nothing but content.

However, if this is the case, it remains unclear how the force can be grasped as anything constitutive of a content's meaning at all: It seems that the force always belongs to the 'hidden side' of a sentence; especially, if Geach's famous recapitulation of the so-called 'Frege Point' is correct, the force would be understood rather as an extrinsic and provisional sort of add-on to the sentence (although a silent one, since Geach believes that the force can only be shown when the sentence is *no longer* being enclosed in a longer sentence) that does not fit to be incorporated into a *Begriffsschrift*. If so, we would not be able to have a sign of force for the explicit indication that the content's truth-condition is satisfied; or, perhaps as the Wittgensteinian 'rule-following' problem revealed, we would have no access to the justificatory conditions of a thought in using our languages after all.

While I agree with the idea that it would be wrong to conceive thoughts as being essentially governed by psychological, stimulative, or other mental regularities—an idea that had been adamantly advocated by Frege—and that thoughts must rather be subject to the logical laws of 'being true' (i.e. of truth-value) and hence be essentially objective, I think it is nonetheless important to maintain the idea that it would be also wrong to regard the objectivity of thought as separable from its 'cognitive value', i.e., from that which is constitutive of an agent's practice of thinking—an idea that had also been proposed by Frege by introducing the notion of sense (*Sinn*). To resolve the tension between force and content, I argue that we should resist treating contents without assertoric force as being logically autonomous and self-contained; rather, we should take Fregean contents to be essentially assertoric.

The Pragmatic and the Semantic

From this point of view, we need to also reconsider the customary contrast between the roles the semantics and the pragmatics deemed to play respectively, which are nowadays often seen as mutually exclusive. Chapter 2, 'The Essential Pragmatics of Assertion', takes a closer look at the meaning-constitutive role of the speaker's pragmatic force. I begin by introducing the problem of rule-following and the problem of context-sensitivity, both of which cannot be properly accounted for if we adopt a Fregean, 'dualist' approach to merely thinkable contents—The dualist approach is best represented by the so-called semantic minimalist view, according to which the pragmatic effects are often understood as limited to optional and additive enrichment that does *not* contribute to the truth-conditional constitution of the proposition and its demarcation—They are, so to speak, considered as being only '*post-propositional*', while the 'pre-propositional' pragmatic effects, if there is any, are rather deemed either trivial (i.e. tantamount to semantics) or reducible to a kind of semantics—Contrary to this common approach, I suggest that there must be a *strong* (i.e., non-trivial and irreducible) yet '*pre-propositional*' pragmatic effect constitutive of any truth-conditional content.

Adopting such a suggestion means that we should not take pragmatics and semantics as always mutually exclusive: We need to see that the speaker's intention is not only constitutive of interpreting what the speaker means by a 'far-side' pragmatic effect (e.g. Gricean implicature) but also constitutive of interpreting what is said—

There should be *no* distinct *two steps* according to which we need to *first* interpret what is literally said through a minimalist semantic process and *then* modulate the 'literal meaning' into an enriched interpretation of what the speaker intended to mean by saying thus and so. To reject the two-step view, and to explain how individual linguistic uses could be constitutive of the meaning of a content, I examine in the dissertation some classic treatments of token-reflexive expressions in literature as well as the debate on essential indexicality (or: essential context-sensitivity). I conclude that a broadly Austinian approach to speech acts is a suitable candidate for us to accommodate an account of strong, pre-propositional view of the speaker's pragmatic force to the Fregean conception of content.

An Act-Theoretic Approach to Assertion

In Chapter 3, 'The Generality of Force–Content Relation', I propose that assertions should be regarded as the *primary species* of all speech acts. Unlike many other speech-act-theoretic variants that treat assertion as one of the many sorts of illocutionary acts (i.e. performative utterances), in this dissertation I propose that we should treat assertion and illocution simply as the same. Such an equation of assertion with illocution is to accommodate the generally Fregean view on the generality of force–content relation, which, as noted, takes contents to be essentially assertoric.

The 4th Chapter, 'Discursive Agency', lays out the fundamental arguments of my approach to assertion that equates the Austinian illocution with the Fregean assertion. Assertions viewed in this way, I argue, must have two basic features: They are *a) commitment-based*—A speaker makes an assertion iff she makes a performative commitment to the discursive propriety (i.e. truth) of the asserted content—and *b) discursive*—To assert is to convey information to the hearer, which, when successful, results in the hearer's integration of the information into his own beliefs.

Under my combined (and modified) Austinian-Fregean view, we can distinguish our acts of thinking or speech fundamentally into two kinds: the full-blooded, committal assertions (which can be seen as 'unqualified' thoughts) on the one hand, and the non-assertoric and thus non-committal, 'distanced' thoughts on the other hand, which correspond to the Austinian locutionary contents secondary to illocutions (thus can be seen as 'qualified' thoughts).

Performative Immediacy and the First-Person

Following the commitment view, I argue further in Chapter 4 that an agent's assertoric commitment also expresses her belief: In performing a commitment, there is no difference between the description of the act of asserting or believing a content and the actual act of asserting or believing that content. The commitment view also allows us to understand the transparency of selfhood, because by making an assertion the agent exemplifies what I call the 'performative immediacy', which implies that the agent always makes commitments in a self-knowing manner. While the agent's commitment signifies the unity of the performative act of instantiating the evaluative norms, whether an act is discursively appropriate (i.e. whether it succeeds in instantiating relevant norms) equally reflects the agent's self-understanding whether its own discursive agency is properly constituted.

If such a characterization of the speaker's performative immediacy is correct, we are then able to conceive assertions as inherently norm-following and knowledgeable performative speech-acts, whose evaluative norms are relative to each speaker's own context. I offer an explanation in my dissertation why the norms for evaluating the propriety of our assertions (the so-called 'constitutive norms' of assertion) must be *a) context-sensitive* and *b) pluralistic*. Yet we can also see that it raises an immediate question whether context-sensitivity could be really compatible with Fregean contents, which are considered to be universally evaluable rather than solipsistic—Both the understanding of selfhood and the understanding of the essential context-sensitivity of content are referring to the same difficulty here: On the one hand, first-person thoughts seem to be, as Frege also admitted, indicating a form of ineliminable context-sensitivity; yet on the other hand, interpreting the ineliminable indexical of the first person as expressing certain privileged access or authority seems to be at odds with the requirement of Fregean contents to be universally evaluable. So how can we understand such ineliminable first person? If, following Anscombe, the first person is not a referential device, so that the selfhood expressed by first-person thoughts must be purely formal and refer to no object in experience (or in Kantian terms, the self must be a purely transcendental subject), how is it still possible for us to comprehend it, especially its ineliminability?

We need to further reflect on the idea of the speaker's first-person authority through a distinction of the argument for 'self-trust' from the problematic argument for

'self-reliance' (a distinction explicitly made by Linda Zagzebski). The argument for self-reliance mistakenly derives the reliability of our first-person thoughts from the reliability of a certain kind of knowledge of ours. The argument for self-trust, by contrast, goes in the reversed order: It derives the reliability or the trustworthiness of certain specific content from the rational commitment of self-trust: When I assert "I have leg pain", I thereby refer to the capacities and know-hows of perceiving, discerning, reporting, memorizing, and bodily awareness, etc. and express trust in these capacities; but it is only possible when they are represented as constitutive of my performative commitment to the discursive appropriateness of my assertions as following the relevant norms in a form of self-trust. Of course, there are more than one thinking agent in the world and more than one context at play in our discourse, and of course disagreements and mistakes could arise in these assertions, but it is not a conflict between self-trust and trust in others; rather, it is a conflict that arises within self-trust. I suggest that only through this could the notion of 'first-person authority' be properly comprehended. I think this also concurs with the Austinian-Fregean view on non-assertoric, distanced thoughts, which are speeches in qualified form derivative of the agent's self-knowing commitment.

Context, Norms, and Disagreement

Chapter 5 'Agency and Constitutive Norms' provides further, more detailed analyses of my account of constitutive discursive agency and discusses its potential controversies and objections. I begin by examining the Kantian constitutivist approach, which contends that there is a separable act of legislation that brings certain norms into being as authoritative. I argue that such Kantian approach cannot be satisfactory. By contrast, I contend that the agent's understanding of the content's discursive appropriateness refers to the conformity to the epistemic standards and norms implicit in the background knowledge of the context of her speech. In other words, demarcating assertions and their truth-values is a question of demarcating the pluralistic background norms at their contexts.

However, we have also seen that we need to explain how we can compare and evaluate each other's performance across different contexts and different background constitutive norms, i.e., we need to answer how disagreements could be possible (i.e. the question of distanced analyzability).

In an ideal conversation, the related evaluative norms are symmetric between the speaker and the hearer, so that the hearer can accept the information frictionlessly and integrate it into his beliefs without the need to abandon or revise any existing belief. (The interlocutors enjoy maximum level of mutual intelligibility and context transparency.) Nonetheless, the norms in a conversation could be conceivably *asymmetric* between its interlocutors—In fact, we find most cases of our discourse as being polemical: Our discourse usually invokes frictions and dissonance; and whenever dissonance (i.e. disagreement) occurs, the interlocutors need to coordinate their background knowledge and hence the norms implicit in their interpretations of the content at stake. Therefore, understanding disagreement and the asymmetry of constitutive norms in a conversation is crucial for us to understand assertion and assertion-evaluation.

Two typical candidate approaches are examined in my dissertation, which I call *(Lewisian) contextualism* and *social constructivism*, respectively. However, I think both are unsatisfactory: The contextualism correctly addresses that whenever we interpret or evaluate an assertion we must take the speaker's own demarcation of background norms into account. But the contextualist strategy insists that to do so we must always preserve the speaker's 'upstream' context, which is obviously expensive and time-consuming. Worse still, it eventually ends up making the assertion-evaluation across contexts and persons impossible—The Lewisian contents, so to speak, cannot be Fregean contents—Moreover, I argue that such feature of contextualism can be compared to the much debated 'faultless disagreement': For a faultless disagreement, the context and hence the constitutive norms of the content have been shifted away from the 'downstream' context of evaluation. The conversation is hence not intrinsically evaluative but exonerative. We must distinguish faultless disagreements from the genuine (the intrinsically evaluative) ones. Unlike faultless disagreements, to genuinely disagree with an assertion is not simply to iterate (fictively) the content at its 'upstream' context; rather, it concerns changing the assertion's original context and its constitutive norms into the context and the norms of 'downstream' evaluation—The social constructivists (e.g. John MacFarlane) typically argue in this direction: Contrary to contextualism, the social constructivist strategy is inexpensive and time-efficient since we no longer need to preserve the speaker's original context and background norms. But while the contextualism hampers the 'downstream' norms in evaluation, I argue that the social constructivism minimalizes the significance of 'upstream' norms. Therefore, I think

both contextualism and social constructivism fail to account for the distinction between faultless and genuine disagreements.

Disagreement and Conversational Dynamics

For a faultless disagreement, the context and hence the relevant norms of the content have been shifted away from the context of evaluation—The discourse is hence exonerative or pathological, e.g. in cases of fictions and mishaps—By contrast, I suggest that in a discourse with genuine disagreement, the constitutive norms must undergo a *dynamic change*: a change from the iteration of the assertion at the 'upstream' context of use (with preservation of the speaker's norms and hence the assertion's appropriateness) to the revision or reevaluation of the assertion at a 'downstream' context (so that the speaker's original demarcation of norms will be no longer considered as appropriate)—The context-shift, therefore, is actually a kind of context-preservation rather than revision—In short, for a genuine disagreement to be possible, the evaluative norm cannot be *simply* external to the hearer—that would be like disagreeing with a drunken man or with a programmed computer—rather, it must aim at the hearer's regret about a mistake and thus eventually at resolution of the disagreement, while according to the dynamic view, our discursive disagreement concerns neither the content's sheer differences in truth-values across different contexts nor the instantiations of its different constitutive norms as such; rather, to evaluate a content it is essential to have its context coordinated, while such coordination is a unique and explanatorily primary dynamic process of 'internalization' of norms that motivates disagreement-resolution and conceptual innovation.

We might regard Robert Stalnaker's idea of proposition-diagonalization as a precursor of such a dynamic view. But unlike his typically metasemantic approach, which exclusively concerns the elimination of possible worlds and reduction of common grounds (while our everyday discourse, by contrast, is usually nonmonotonic), I suggest that we must have a fundamentally normative approach to the dynamics of change: As indicated, our major challenge consists in the task of accounting for the asymmetry of norms in discourse—In such a discourse, the 'asymmetrical' norms are in a sense not implied, yet in a sense indeed implied, in the hearer's background knowledge—In Chapter 6, 'Dynamics of the Discursive Power', through examination of some fundamental ideas of the dynamics of belief revision, I argue that the most

important consequence of the ideas of doxastic and conversational dynamics is that preserving a context and revising it are not two steps, neither are they two separate operations; rather, they must be seen as two aspects of the same dynamic process—while context-revision is explanatorily primary, context-preservation needs also to be incorporated as a constitutive part into the explanation of the genuine disagreement that aims at the hearer's acknowledgement of mistake—Accurately speaking, the genuine disagreement aims at the hearer's recognition of a norm as *counterfactually* constitutive of his previous assertion, which renders that assertion neither 'simply' false nor exonerated but having a unique and irreducible normative feature of 'regretfulness'. In my dissertation, I indicate that such an irreducibly normative feature can be found in our linguistic expressions with the so-called 'backtracking counterfactual' phenomenon.

The Rational Power in Assertion

The most important point of the dynamic view, however, is that although the evaluation of a content's truth or propriety is context-sensitive and hence dependent on the particular agent's background knowledge, the context or the particular agent cannot be seen as an ineliminable feature of the content. This amounts to the view that we must resist reading assertions simply as a kind of function, the interpretation of which yields truth-values at different contexts. At the end of the dissertation, I offer some reflection on the idea of the power of thinking (or equally, the power of being engaged in rational discourse) and suggest that we should also resist a functional reading of the power of thinking. On the contrary, the manifestation of our discursive power of thinking should be comprehended precisely in terms of the dynamic movement of internalization of norms rather than the norms' sheer instantiation as such. Only in this dynamic way can we understand how our assertions, although their constitutive norms are essentially pluralistic, socio-historically relativized, and polemic, can nonetheless be seen as manifestation of a form of power. It is important to see, moreover, that the dynamic picture does not offer a meta-level functional generalization of the norms for thinking; rather, it indicates that our practice of thinking cannot be exhausted by explicit linguistic instantiations of a specific set of given norms supposedly constitutive of the practice but always asks for further discourse on implicit disagreement and its resolution.

CHAPTER 1

Frege's Assertoric Force

1.1 Meaning and Use, Semantic and Pragmatic

What distinguishes thoughts from other occurrences and constituents of our mental life, from sensations, impulses, inclinations, and the rest? This question typically belongs to those age-old problems of philosophy. An instant and natural reply to the question is that thoughts are intelligible contents, and moreover, they are intellectually sharable. Unlike a muscle twitch, a murmur, or any other unintelligible expressions such as "abracadabra", the expression of a thought is considered meaningful. Should we accept such a description of thought—and I think there is no good reason we should not—our question is then a question about the condition of demarcating meanings. Let us call this question the *question of demarcation*.[1]

It thus seems that the question of demarcation asks for a theory of meaning. At the beginning of his paper "Identity and Predication", Gareth Evans made the following claim:

> A translation is one thing, a theory of meaning another. A manual of translation aims to provide, for each sentence of the language under study, a way of arriving at a quoted sentence of another language which has the same meaning. A theory of meaning, on the other hand, entails, for each sentence of the language under study, a statement of what it means. A translator states no semantical truths at all, nor has he any need of the concepts of truth, denotation, and satisfaction. Semantical truths relate expressions to the world, and can be stated only by using, not mentioning, expressions of some language or other. (Evans 1985a, 25)

Thoughts need to be expressed in languages. Yet the occurrence of language alone does not suffice to demarcate meanings. Evans draws the distinction between *mentioning* and *using* expressions of a language, and points out that merely mentioning an expression

[1] Cf. Dummett (1993a, 154).

may not involve any understanding of the meaning of that expression. He compares the mere mentioning of an expression to the translation according to a manual. This is a spot-on comparison. While machine translation is a commonplace nowadays, hitherto there is still no good reason to regard computers as having the same kind of intelligence as we homo sapiens do. And the notion of meaning seems to be one of those typically philosophical concepts that cling on to our perceived difference from the machines.

A fervent fan of AI might maintain that our difference from the artificial neural networks is rather quantitative rather than essential, so that our insistence on the significance of meaning is no more than illusory. The problem, however, is that we only adopt such a view at the cost of dispensing with too many notions central to our mental life (meaning, essence, rationality, contradiction, truth, … the list of terms could be indefinitely long); and more seriously still, since such an approach cannot eliminate all the 'philosophical' notions, it rather narrows them down to a very limited set of behavioristic terms. And these terms fall short of offering a satisfactory account of the basic features of our thinking.

If we cannot dispense with meaning, we need to better understand it. And Evans's proposal sets us a good example. A machine can translate an input English word, say, 'house', into German and outputs 'Haus'; or it can return a lexical definition of the input word, e.g. "a building for human habitation". In either case, we can regard the outcome as made according to a 'manual' (or database). In other words, it exemplifies a connection between the mentioning of two words, or between the mentioning of a word and a string of words. But a connection like this does not in itself exemplify any understanding of the meaning of language. (Although it could serve as material or a non-independent part of our intellectual life.) In sum, merely mentioning an expression in a language is not yet *using* the language and hence does not suffice to demarcate meanings.

Using a language, by contrast, requires performance of an intellectually significant act, i.e., an act of thinking. The distinction between mentioning and using indicates that the criteria for the demarcation of meanings do not just consist in the apparent linguistic features—Both "abracadabra" and "Caesar died" can be generated with the same probability by a computer as a random string of 11 characters. Though the latter has an apparently standard syntactical structure, it *in itself* does not yet qualify as a meaningful expression, because it failed to represent an intellectually significant act of thinking. Consider, furthermore, that a tired lab scientist might have mechanically

written down all the randomly generated strings he observed on the computer screen without ever noticed "Caesar died" to be of any difference from other strings. The phrase thus turns meaningful only when the scientist later discovered it and reported it to his colleague: "Can you believe it? The computer said Caesar died!" It is meaningful not because the computer 'said' it—In a sense the computer did not 'say' anything—The scientist, on the other hand, is not just 'quoting' a speech. It is rather the scientist's act of speech that 'disquotes' and thus demarcates the phrase as meaningful. To use Wittgenstein's words, in this case only the scientist made a move in the language game.

At the beginning in *The Blue Book* Ludwig Wittgenstein points out that in order to understand what is the meaning of a word, or what is the meaning of 'meaning', we should better start by comprehending what is it to *explain* the meaning of a word. I interpret Wittgenstein's proposal as follows: An expression has meaning only if it is answerable to evaluations of its appropriateness to serve as an explanation. To establish a theory of meaning, therefore, we should begin with investigation into the nature of the act of thinking, while the nature of the act of thinking consists in the language-user's practice of explanation. In short, it is the explanatory act (*Deutung*) that makes meaning (*Bedeutung*) possible—In other words, it is the pragmatic force of a speech-act of explanation that makes a sentence semantically 'alive'.

To regard a meaningful expression as a speech act, therefore, is to emphasize the significance of pragmatics in determining meanings. It thus does not suffice to only discern the 'literal', lexical meaning of a phrase to demarcate its meaning—That would be like referring to a translation manual, which, as indicated, does not necessarily involve the understanding of meaning—Instead, to comprehend and evaluate an expression as a speech act we must know a great deal about the speaker's intentions and epistemic implicatures (about evidence, credence, expertise, etc.) To use Donald Davidson's example: It should never be a naïve thing to assert that the price of plutonium is rising.[2]

Yet, there is another crucial feature of our language: we express infinite possible meanings and concepts using only a finite number of phrases and ways of their combinations. Even though it is up to the speakers' intentions to say what they mean under different circumstances and at different contexts, once the phrases are said, their meanings are not tied to those circumstances and contexts. Hence, even though the

[2] See Davidson (1982, 183).

pragmatics plays an importance role in our languages, it is nevertheless commonly suspected that the pragmatics cannot be central (but rather, say, secondary) to the question of meaning-demarcation. The skeptics would argue, for example, that if I intend to mean 'mouse' but have said 'house' instead, it is rather the latter—namely, that which is said, i.e. 'house'—that plays the central role in the identification and the sharing of meanings.

The feature of our language shows that meanings are not entities that belong to the individuals' 'private episodes' that take place at particular points in space and time. Quite the contrary, they are sharable contents that can be objectively expressed and understood among different individuals at different times. Therefore, to some, the emphasis of pragmatics for a solution to the question of demarcating meanings might sound like a dangerous drift into lawless wilderness of speech—It might seem, so to speak, that that would allow everyone's own arbitrary will to mean (*deuten*) what they say so that the semantic meanings can be never stabilized and hence identified.

Indeed, this is a legitimate worry. For a speech act to be meaningful at all, its content must be generally intelligible. The question of demarcation is inherently a question of disquotation, and it hence makes no sense to talk of a language-user's pragmatic force if what he speaks is a private language. The problem of a private language is that we could not find any 'anchor point' for mutual intelligibility while using it for communication—Of course, I can use the English language innovatively. I could decide to utter: "I have stomach mumbo jumbo" to express the concept that I have stomachache. But it is only intelligible when I follow at least part of the conventions of English—For me to convey a sensation, I can only express it in terms of generally comprehensible perceptual features graspable by concepts—If I fail to do so, the problem is not that you cannot feel the pain in my stomach, but that when I pronounce the word "mumbo jumbo" you would have no idea what kind of meaning I intend to convey. For this reason, Evans claims that a language-user states 'semantical truths'. Uttering an expression in this sense is also passing a *judgment* on the semantical truths of the expression. (It might be controversial what 'semantical truths' are, but it should not be controversial here to say at least that meanings must be semantically stable and publically sharable.)

In the opening section of Frege's *Logical Investigations* that he undertook in his final years but did not finish, he addressed the importance of grasping the *thought* (i.e. the universally sharable intellectual contents) as independent of any of our

'representational' mental activities, which he takes to be the necessary starting point of any inquiry into logic and meaning:

> Not everything is a representation (*Vorstellung*). Thus I can also recognize thoughts as independent of me; and other people can grasp them just as I do […] We do not *have* a thought as we have, say, a sense-impression […] So it is advisable to choose a special expression here; the word 'grasp' (*fassen*) suggests itself for the purpose. To the grasping of thoughts there must then correspond a special capacity of mind, the power of thinking (*Denkkraft*). In thinking we do not produce thoughts, we grasp them. For what I have called thoughts stand in the closest connection with the truth. What I recognize as true, I judge it to be true fully independently of my recognition of its truth and also independently of my thinking of it. (Frege 1918–19, 74)

This is an attractive idea. By emphasizing the independence of semantic truths from any particular person's representational act, Frege believes we can thereby maintain the objectivity of meaning, so that it is essentially an issue not that by uttering *φ someone intends* to mean *F* but that the utterance of *φ*, *in itself*, means that *F*.[3]

The most straightforward approach to the independency of semantic truths is to claim that each word or each singular term that we use has a determinate, literal meaning. Whenever we use a singular term it rigidly designates and hence picks out a unique object or a unique type of objects in reality; and a thought is demarcated if and only if an object or type of objects is picked out and thus spoken of. Call this approach *simple referentialism*.

However, such a referentialist emphasis of a linguistic expression's semantic truths as its immanent property has serious difficulties. As a famous example made by

[3] Specifically, this passage of Frege's suggests two ideas: First, as contents of judgments and logical thinking in general, our thoughts have nothing to do with contents of our material life (e.g. the sensory contents). Such an approach thus drastically diverges from Kant's approach in his transcendental logic, which is considered to provide a philosophical foundation for our empirical knowledge at large. This divergence can be clearly seen in Frege's use of the term 'representation' ('*Vorstellung*'), which means nothing objective but those that are rather qualia-like. (Discussion on Frege's and Kant's approaches to logic cf. (MacFarlane 2002).) Second, even though here Frege suggests that we need to have a special mental capacity (the '*Denkkraft*') for thinking—namely the 'grasp' of thoughts—the mental capacity itself nonetheless does not serve the role of demarcating meanings and thoughts. But for this very reason it remains unclear whether the Denkkraft has anything to do with Frege's theory of meaning (or whether it is just another psychological mechanism of our mental life that needs to be purged from his theory of meaning).

Frege, 'Hesperus', 'Phosphorus', and 'Venus' are considered to have the same semantic truth, i.e. reference to the planet Venus; yet in our conversations, 'Hesperus is Hesperus' seems trivial, while 'Hesperus is Phosphorus' and 'Hesperus is Venus' seem informative. Furthermore, they appear to be not *salva veritate* substitutable in some contexts: E.g. the sentence "Phosphorus is the brightest object in the sky after the Moon before sunrise." remains true if we substitute 'Venus' for 'Phosphorus', but it will be false if we substitute 'Hesperus'.

Different approaches can be made at this point to revise the theory of meaning. Very roughly, we can distinguish two kinds of approaches:

The first kind is to hold fast to the idea that thoughts are semantically autonomous—Namely, a thought is nonetheless a self-contained unity with its semantic truth intact, regardless of its different uses in different contexts—Its major difference from simple referentialism is the re-conception of the *unity* of meaning: For simple referentialism, the basic semantically autonomous unity can be a word or a phrase. But the revised view contends that the basic unity must be a *sentence* or a *proposition*, which, in general, manifests a predicative structure. It thus explains that the falsehood of the sentence "Hesperus is the brightest object in the sky before sunrise" is not because any singular word in that sentence fails to have a meaning but because the sentence as a whole fails to exemplify a meaningful predicative relation. In other words, for the first kind of approach the demarcation of a thought still consists in picking out the object the thought is referring to; but unlike the simple referentialism, according to which the reference can be achieved through the correspondence between a word and a thing, the revised approach contends that the reference is achieved through the correspondence between the determinate meaning of a complete proposition and a fact described by that proposition. (Cf. the famous opening statement of Wittgenstein's *Tractatus* that the world is the totality of facts, not of things.)

The second approach to revise the referentialist theory of meaning is to maintain the significance of pragmatic force of use in the demarcation and determination of meaning of a content at a particular context—Similar to the first kind of approach, this second approach acknowledges that the unity of meaning primarily consists in a whole sentence or proposition instead of singular words or phrases. But unlike the first approach, according to which such unities of meaning are semantically self-contained

and thus are pragmatically either irrelevant or only minimalistically relevant,[4] the second approach holds that the demarcation of meaning essentially consists in the pragmatic force of use. In other words, we understand the unity of a proposition or a thought only in terms of its corresponding *unifying act of utterance*, not vice versa. The second approach therefore rejects the idea that a thought could be semantically self-contained and independent of the acts of uttering it; instead, it contends that the solution to the demarcation question is essentially an act-theoretic and pragmatic one.

Which of the two approaches did Frege adopt? In some sense, the appropriate answer should be: Both. The distinction above is very broad-brushed, and an insistence solely on either of the two approaches could lead to both advantages and disadvantages. The advantage of the first approach is that it straightforwardly guarantees the semantic stability of thoughts, and it sets a very convenient foundation for structural analysis of compound thoughts; yet its disadvantage is that it neglects the use-sensitivity of contents in contexts. (In fact, as will be argued in this dissertation, it makes it almost impossible to explain the use-sensitivity.) The second approach, by contrast, has the advantage of accounting for the contents' use-sensitivity; yet it might fall short of guaranteeing their semantic stability across contexts and circumstances of use, and hence risk drifting into the lawless wilderness of speech. It hence constitutes a lasting challenge to meaning-theorists to reconcile the two approaches and avoid their relative disadvantages.

Frege's own effort to reconcile the two sides, as is well-known, is to distinguish between the reference (*Bedeutung*) of a content on the one hand, and the sense (*Sinn*) of the content as a way of thinking its reference (i.e. as a mode of presentation) on the other. However, later I will argue that such distinction is at least confusing, if not misleading, to the problem of meaning-demarcation. In general terms, Frege's failure to balance the two sides roots in an obstinate misconception of the distinctive roles that pragmatics and semantics play respectively: According to such a distinction, what the pragmatics is about only belongs to one's particular uses, such as her personal linguistic competence and verbal preferences, etc., while the semantics accounts for the universality of thought, namely those which remain the same regardless of their particular uses. If we stick to such a distinction, it would be no wonder that the pragmatic effect of an individual speech-act is excluded from being accountable for

[4] More discussions on minimalism see §2.2.

demarcating its content as meaningful.

In fact, the success of the Fregean analyses in formal logics has reinforced such distinction, and it makes the second kind of approach seem unlikely to succeed. The common suspicion is precisely that the pragmatic effect of a speech is too 'exotic'. (I speak of 'exotic' in the sense that the pragmatics functions only as a 'post-propositional', supplementary and perhaps arbitrary enrichment of the meaning of a proposition determined by semantics alone.) For example, Richard Gaskin once argued:

> […] It might be tempting to suppose that what is required to turn a bunch of words into a sentence is the right kind of linguistic intention, that, for example, the words 'Socrates is wise' will constitute a sentence on any particular occasion of use, as opposed to a mere list, if they are uttered or inscribed with the right intention. […] Of course, if by 'use' we mean not use on any particular occasion but something tantamount to semantic meaning ('how the word is used in the language'), then it would be correct, but unhelpful, to say that use is what unifies a sentence: the question will then be what that unifying use consists in. And to that question only a reply that specifies the semantics of sentences in a way that casts light on their unity will be satisfactory. (Gaskin 2008, 390f.)

Not everything in our mental life is a practice of thinking or intention. Therefore, for the second kind of approach it is crucial to maintain that the pragmatic force is not any sentient behavior or stimulus-driven mechanism of our material life; instead, it must be *sui generis* and reserved for the use of language. (Recall Frege's notion of the so-called 'Denkkraft'.) However, according to Gaskin, if the speaker's intention is not at odds with the semantic truth but considered to account for the unity of a proposition and its truth-conditions, it must be trivial since it can be nothing but tantamount to the unity of meaning originally sponsored by semantics alone.

The main task of this dissertation is to defend a solution to the demarcation question generally in line with the second kind of approach to the theory of meaning. In so doing, we need to resist distinguishing between the pragmatic and the semantic roles of speech according to a contrast between user-arbitrariness and objective stability; moreover, to respond to the challenge to the pragmatic approach, we need to better understand what a unifying act of use consists in, if it is neither tantamount to the semantics of the proposition nor at odds with it. I will argue that the pragmatic force is not only not at odds with the semantic stability of contents, but also is the key to understanding it. (Thus, I reject the idea of the first approach to meaning that a thought

could be semantically self-contained and independent of the acts of use.) By giving a speech-act-theoretic approach to meaning-demarcation and the unity of proposition, I contend that the unity of thought (perhaps with the exception of mathematical contents and objects) essentially consists in the 'pragmatic unity' of the discursive act of thinking and asserting.

In fact, as I will discuss later, this idea reflects Frege's own original insight into the assertoric force. (But due to his insistence on the contents' semantic autonomy, his insight into the force eventually leads to obscurity.)

1.2 Frege on the Content and the Force of Assertion

Frege's original insight into the force indicates that the problem of meaning is a problem of the condition for the 'meet of minds'. The idea that thoughts must be publically sharable has been consistent throughout all his works. When we talk to each other, we are communicating mutually intelligible information. We may speak different languages and have distinct culture backgrounds, but we nevertheless regard our interlocutors as fellow human beings. We may distinguish a member of any human-language speaking group regardless whether she is English or Inuit or Swahili from a non-human being according to a unitary thinking ability, which is considered as the power of conferring thoughts.

According to Frege, what demarcates a meaningful content from other things is that it consists of semantic truths, and to demarcate such a content is to pass a judgment on its truth—This is precisely what Frege has proposed in his *Begriffsschrift* (1879). Therefore, it would be good to begin our investigation into the question of meaning-demarcation and the unity of proposition with Frege's envisaged project of *Begriffsschrift*.

1.2.1 The Concept Script

To see why the semantic truths are constitutive of expressions of meaning, consider a conversation between Sophie, a scientist, and Jack, who does not believe in climate change. Sophie tries to convince Jack that there is strong scientific evidence of climate change. Now, after Sophie elaborated her argument, Jack speaks: "Look, I appreciate what you've explained to me. You seem to have very good reason to believe what you

believe. That's fine, but I'm not convinced. I'm not familiar with whatever you scientists have to say, but at least I believe that everyone is entitled to their own opinion." "But climate change is not an opinion, it's a fact!" Sophie protests. "Still, this is *your* opinion." Jack replies.

One may find Jack's seemingly reconciling, 'non-disputandum' attitude in the conversation actually irritating. Why is it so? Doesn't the appeal to the idea that everyone is entitled to their own opinion represent a virtue of tolerance and democracy? The reason for the irritation is precisely that expressions of meaning are constituted of judgments, which are claims of fact. In the example above, when Sophie claims that *P*, she acknowledges *P* as her belief. But holding a belief is more than just being in a certain mental state, and expressing it is more than declaring or revealing one's mental state. That is why Sophie's claim is not just a mentioning of *P*, but a judgment, i.e. she states *P* as a fact. In doing so, she must be able to *explain* it to show that her believing in *P* is justified—When needed, she must be able to give further reasons, such as *Q*, to support her belief *P*.

The difficulty, however, is that when someone asks Sophie: "why do you think *P*?", even though she may respond him by referring to *Q*, yet her interlocutor may still ask: "why do you think *Q*?"—It is a legitimate question, since *Q* is also Sophie's belief and thus needs to be grounded, too. Obviously, questions of "Why?" can go on and on and on, nonstop, regardless whether she cites *R* as a further reason for her belief, or cite a higher-order belief about *Q*, i.e., the belief that *Q* is a reason for her belief *P*. Whatever Sophie proposes, it is expressed in form of a belief. Since, for Jack, no belief can be taken as a fact without qualification, the infinite regress is unavoidable. It might thus turn out that Jack's response is not so unreasonable after all, namely, what everyone has can be nothing more than his or her own opinions. Or is it?

We may see that Jack's response is arguing in the exact same way as the Agrippan skeptics did.[5] As a result of such skepticism, no judgment and hence no knowledge could be possible. Such result concludes that we can only talk by mentioning the strings of names and words—in other words, always in quotation—but never in meaningful thoughts: When I utter a token of sentence "*P*" and claim that "*P*" is true, I could only be claiming that I have certain positive attitude expressible by "*P*", while you may well have a negative attitude expressible by the same token you hear.

[5] On the Agrippan skepticism cf. Bett (2005), cf. also Franks (2005), Kern (2017).

However, if this should be the case, then we can never grasp the meaning of a sentence, since it would be impossible to determine whether the contents expressed by the two sentence-tokens actually refer to a same sentence type or not—according to the Agrippan skepticism, such a determination can only be carried out as a further token of expression, which is itself groundless.[6] Then all our utterances can never be disquoted away, and everyone would speak in his or her own private language, and any meaningful conversation would be impossible, too.

When Sophie claimed that climate change is a fact, Jack responded by saying that "you only *think* that climate change is a fact". But such procedure ends in a hopeless skepticism. His response is irritating because he frustrates his interlocutor by disqualifying her claim of fact, reducing it to nothing more than just an opinion. The disqualification is a high-order move, which is a radical refusal of first-order reasoning. But the skeptical result indicates that the tolerance of such refusal itself cannot be reasonably justified, too: If "… is a fact" is the form of the expression of knowledge, then the attempt to reject this form cannot maintain consistent by itself— To use Grice's words, in such case Jack himself actually violates the 'cooperative principle' necessary for a conversation to take place. (We may also see that Descartes has employed the same approach in the well-known argument in his *Meditations*.)

Now, as mentioned, Frege has also put persistent and tremendous effort throughout his philosophical career into the work of establishing a universal conceptual scheme to dispel the skeptical hopelessness. For him, a thought is a *conceptual content* which can be objectively grasped by any rational human being. The language of such objective universal conceptual content is to be written in the 'concept script'—a 'formula language of pure thinking'. His first major work, the *Begriffsschrift* (i.e. the

[6] It might be pointed out, furthermore, that in this case we can even not determine whether the physical occurrences of two tokens of locutions represent the *same* type of locution or not. In other words, if the claim of fact can be nothing more than mere opinion, then not only there would be no fact, but also there might be no opinion—And what we have would be only private immediate mental occurrences that can never be grasped and conceptualized at all. James Conant suggests that while we may call the impossibility of fact the *Cartesian* version of skepticism, the impossibility of opinion or belief can be ascribed to the so-called *Kantian* version of skepticism. Cf. Conant (2004).

Concept Script), which announced the beginning of modern logic, is explicitly dedicated to such a project.[7]

For Frege, the basic unit for such a Concept Script is a sentence, i.e., a proposition. It has meaning by expressing truth, i.e. fact. For us to determine its meaning is to judge its truth. As early as in *§3* of his *Begriffsschrift*, immediately after putting forward the notion of judgment, Frege emphasizes that the subsentential grammatical structure of a sentence has nothing to do with the meaning or truth it expresses. His example is the following two sentences: "The Greeks defeated the Persians at Plataea", and "The Persians were defeated by the Greeks at Plataea." These two sentences, though having different subjects and predicates, intuitively express an identical judgment about a same fact. Hence, Frege claims that the conceptual content is that which remains the same in both sentences, and declares that "only the conceptual content is of significance for his *Begriffsschrift*".[8]

The two sentences above express the same conceptual content. For Frege, what makes them different is irrelevant for the inquiry of judgment and debarred from their conceptual content. He calls the parts that do not belong to the conceptual content the mere 'wording' (*Wortlaut*) or the 'coloring' (*Färbung*) of expression. The conventional differentiation between subject and predicate in a sentence hence belongs to the wording, because, for Frege, "[i]n language the place occupied by the subject in the word-order […] is where we put what we want the hearer to attend to specially".[9] The subject–predicate distinction, in other words, only remains to have rhetoric importance. By contrast, the conceptual content should hence be expressed in a form of 'colorless' script, i.e. the Concept Script. Its idea refers back to Leibniz's project of '*characteristica universalis*', i.e. an artificial language of science purged of wording and coloring, which can eventually overcome the confusion of tongues and the babel of our natural languages.[10]

---

[7] In what follows, I will use '*Begriffsschrift*' to denote Frege's book, and 'Concept Script' to denote the idea of the language of pure thinking that Frege endeavored to undertake.

[8] See Frege (1879), §3.

[9] Ibd.

[10] One famous example is that the words 'and' and 'but' are treated by Frege as equal at the logical level. Cf. Frege (1879, §7), and idem (1918–19, 64).

The relation of a conceptual content to its wording and coloring, therefore, seems to be like the relation of a pizza to its toppings and condiments. But the wording and coloring may be not only supplementary but rather necessary to the content—In a sense, a sentence is materially composed of nothing but its subsentential components or features, e.g. a bunch of words or writing characters or string of phonetic syllables, etc.—However, the key here is to view the content as an explanatorily prior unity—Thus a sentence, in another sense (i.e. with respect to its 'form' as a conceptual content) is more than any of those subsentential components—Frege enunciated this priority explicitly:

> [I]nstead of putting a judgment together out of an individual as subject and an already previously formed concept as predicate, we do the opposite and arrive at a concept by splitting up the content of a possible judgment.[11]

This idea is central to Frege's unique and revolutionary attempt to offer a universal, logical formalization of the *assertoric force*—Before Frege, people had already been practicing the designation of certain form of general univocal concept preferably written in uniformed scripts: We may simply stipulate a set of conceived scripts the meanings of which remain the same in all their interpretations, e.g. using schematic letters as we please: '*a*', '*b*', '*c*'… or '1', '2', '3'… —just like the customary natural numbers in mathematics.[12] But the issue becomes tricky if we want to *explain* our acts of applying and understanding such stipulated scripts as universally intelligible and hence equally formulate this very explanation as part of the script as well. Frege, who understands his work as a clarification of the foundation of arithmetic, undertakes to explain this using the idea of the assertoric force: since the force designates the act of judging or asserting

[11] Frege. "Boole's logical calculus and the *Begriffsschrift*". In Frege (1979, 17). Cf. also Robert Brandom: "there is available a sort of answer to the question, 'What are sentences, and why are there any?' that is not available for any subsentential expression […] Without expressions of this category there can be no speech acts of any kind, and hence no specifically linguistic practice. " In idem (1994, 363). Note that, strictly speaking, in Frege's technical terms the concept itself is not a subsentential component but the concept-word (Begriffswort) is. (The concept is rather the referent of a concept-word.)—For Frege, concepts and contents differ from concept-words and sentences because for the first two the 'sensory clothing' of our language has been removed—I shall briefly discuss Frege's use of the notion 'concept' in §1.4.3. As I will indicate, as long as the problem of meaning-demarcation is concerned, the tension between repeatable sensical linguistic components and general truths persists.

[12] The notion 'schematic letter' is used by W. V. O. Quine to designate the symbols that are not bound by a quantifier, in contrast to the variables, which are bound by a quantifier.

a content to be true, it demarcates a content *as a content* in the Concept Script. Moreover, since meaning should only be demarcated by what is meaningful, the Concept Script must be conceptually self-explaining—In other words, the assertoric force must also be able to be written in the Concept Script.[13]

In *Begriffsschrift*, Frege invented the famous (and notorious) *assertion-sign* (⊢) for the force. The symbol was adopted in *Principia Mathematica*,[14] and later became the 'turnstile' notation for syntactic consequence in mathematical logic. But accurately speaking, in Frege's *Begriffsschrift* the sign originally consists of a vertical stroke (|) combined with a long horizontal one (—): While the vertical stroke is called by him the *judgment-stroke* (*Urteilsstrich*), the horizontal stroke is called the *content-stroke* (*Inhaltsstrich*). Hence, the formula:

(1) — *p*

expresses a content of possible judgment, which, however, is not yet asserted. And only when the judgment-stroke is added to (1), i.e.:

(2) |— *p*

is the judgable (i.e. 'assertable') content converted into a judgment. Formula (2) hence expresses a judgment that *p is the case* (i.e. *is true, is a fact*).

Therefore, we might want to say that only the vertical stroke designates the assertoric force, not the whole assertion-sign (⊢). But as will be discussed later, the distinction between the judgment-stroke and the content-stroke is indeed confusing—

[13] Cf. "On Mr. Peano's Conceptual Notation and My Own", where Frege emphasizes the importance of his invention of the notation for the assertoric force in his *Begriffsschrift*:

> "[…] I have introduced a special sign with assertoric force, the judgment-stroke. This is a manifestation of my endeavor to have every objective distinction reflected in symbolism. With this judgment-stroke I *close off a sentence*, so that each condition necessary for its holding is also effectively to be found within it; and by means of this *self-same sign* I assert the content of the sentence thus closed off as true. Mr. Peano has no such sign […F]or Mr. Peano it is impossible to write down a sentence which does not occur as part of another sentence without putting it forward as true." (Frege 1984, 247f. My emphases).

[14] "In ordinary written language a sentence contained between full stops denotes an asserted proposition, and if it id false the book is in error. The sign '⊢' prefixed to a proposition serves this same purpose […] Also a proposition stated in symbols without this sign '⊢' prefixed is not asserted, and is merely put forward for consideration, or as a subordinate part of an asserted proposition." (Whitehead and Russell 1963, 8).

and from mathematical logical perspective, a failure. Nonetheless, I will suggest that we can observe some interesting philosophical concerns in Frege's original usage of the notation.

1.2.2 Frege's Original Insight

As indicated, in his *Begriffsschrift* Frege rejects taking the grammatical subject–predicate distinction as having any primary significance in the theory of meaning. He turns the focus to a force–content distinction instead.[15] According to the force–content view, we should view the unitary whole of a sentence as a conceptual content, whose explanatorily primary unity is demarcated by the assertoric force. Such a radical move Frege made in his *Begriffsschrift* brings us two equally important theses—we may call it Frege's (Original) Insight: *(a) the assertoric force is meaning-constitutive; and (b) the force cannot be assimilated to a predicate*.

Ad (*a*): The assertoric force indicates the act of assertion or judgment that determines the truths of a content, i.e. the act that asserts that a conceptual content refers to a fact. Therefore, the assertoric force is constitutive of the meaning of a conceptual content, and it demarcates a content as a meaningful unitary whole.[16]

Ad (*b*): Even though the force is meaning-constitutive and belongs to the Concept Script, and even though it demarcates the content's unity, it does not contribute to the content. It means that the force is not assimilable to a kind of content, i.e. as if saying that the last brick that completes a wall is nonetheless a kind of brick—Even though the assertoric force is said to convert a content *p* into an assertion that says that *p* is true, the expression '… is true' hence is not a real predicate and makes no

---

[15] Peter Geach acclaims this insight of Frege as a great logical achievement: "It took the genius of the young Frege to dissolve the monstrous and unholy union that previous logicians had made between the import of the predicate and the assertoric force of a proposition. Even when a proposition has assertoric force, this attaches to the proposition as a whole, not specially to the subject, or to the predicate, or to any part of a proposition." (Geach 1980, 51).

To be fair, subject–predicate analyses are not always purely grammatical but have metaphysical bearing. But Frege's worry, as I understand, is that the subject–predicate analyses seem too easily to couch themselves in the 'traditional' metaphysics of classifying things and properties instead of reflecting upon concepts and their object-relations. Similar attack on subject–predicate analyses cf. Strawson (1959), §5.1.

[16] As will be discussed in more detail later, this means that the content-stroke will need to be viewed as derivative of the judgment-stroke.

contribution nor any alteration to the content it is attached to—As Michael Dummett summarizes:

> The judgment-stroke is the sign of assertion proper, that which carries the assertoric force. It is therefore not a functional expression, or part of one: we cannot enquire of it what its sense is, or what its reference is; it contributes to the meaning of the complex sentential symbol in quite a different way. Furthermore, it is only the sentence to which the judgment-stroke is prefixed which may be said to express a sense or to stand for a truth-value. (Dummett 1981, 315f.)

For Frege, the truth condition of the sentence '*p*' and that of the sentence '*p* is true' are just the same, and they express the same thought. It means, moreover, that the force should not be viewed as certain conceptually inscrutable and additive ingredient that turns the content into something other than such, too—It is thus not only inappropriate to compare the force–content relation to how a special kind of brick is to other bricks, but also inappropriate to compare it to how the wall is to the bricks. (This means that the conceptual contents, too, cannot be assimilated to words, phrases, or any other subsentential components.)

There are many potential challenges to Frege's Insight that one needs to respond to. We will have more discussions of them later in this dissertation—Besides, although I support Frege's Insight, I will also argue later that Frege himself did not live up to it—There seems to be, above all, a tension here: on one hand the force is claimed to be demarcating a content as complete; but on the other it is said that the force is neither additive nor assimilable to the content, and the content is claimed to be already complete.[17] So how should we conceive of it after all? It is indeed a tricky question. But before we dive into further exploration of it, let us make a little effort to clarify some terminological complexities:

For Frege, the introduction of the judgment-stroke is central to his whole logical project. In a 1906 manuscript of *Introduction to Logic*, Frege begins, in a section titled "Dissociating assertoric force from the predicate" (*Die Ablösung der behauptenden Kraft vom Prädikate*), with the following passage:

> We can express a thought without asserting it. But there is no word or sign in language whose task is simply to assert something. This is why, apparently even in logical works, predicating is confused with judging. As a result one is never quite sure whether what

[17] I will make a more detailed discussion of the force–content relation in Chapter 3.

> logicians call a judgment is meant to be a thought alone or one accompanied by the judgment that it is true. […] *I use the word 'thought' in roughly the same way as logicians use 'judgment'*. To think is to grasp a thought. [*Denken ist Gedankenfassen*] Once we have grasped a thought, we can recognize it as true (*make a judgment*) and give expression to our recognition of its truth (*make an assertion*).[18]

In the first part of this quoted passage Frege emphasizes that the act of judgment or assertion is not an act of predication, since, according to his Original Insight, to say something is true is not predicating any extra property to that thing. The second part of the passage is yet somewhat puzzling, since it seemingly distinguishes three different types of acts: grasping a thought (i.e. thinking), judging, and asserting. In this dissertation, however, I will take the two notions 'judgment' and 'assertion' to be synonymous—*Pace* the subtle difference between the judgment-stroke and the assertion-sign, both 'judgment' and 'assertion' indicates the content equipped with the assertoric force—Apparently, Frege considers the act of judging to be taking place in the 'mental sphere', while the act of asserting rather denotes the phonetically expressive acts of utterance. (This can be reflected by his adoption of the concept of *characteristica universalis* as a form of written rather than spoken language.) But I think such a behavioral differentiation between judging and asserting is rather trivial to our investigation. Furthermore, even though a judgment can be more adequately deemed as being made silently (while for an assertion less adequately so), we should notice that such an act is nevertheless *inherently discursive*—as I will argue in more detail later.[19]

The second terminological issue concerns notions of the content: In this dissertation, I take 'conceptual content', 'thought', 'proposition', and 'propositional content' to be all synonymous. (If without specification the term 'content' is also used as short for 'conceptual content'.) But we need to be a bit careful with the latter two

---

[18] Frege (1979, 185). In a short (only 10-line long), fragmentary note supposedly prefatory to the *Introduction to Logic*, Frege tries to answer the question *"What may I regard as the Result of my Work?"* by summarizing the characteristics of his logic that he considers important, which begins with the declaration: "It is almost all tied up with the concept-script". After enumerating some of its features, the line breaks off, and a new line starts: "strictly I should have begun by mentioning the judgment-stroke, the dissociation of assertoric force from the predicate…" Ibd. p.184.

[19] In what follows, if without direct association with certain quoted passages (e.g. Frege's), I will generally stick to the use of the notion of 'assertion' instead of 'judgment'.

here: It is commonly considered that 'propositions' or 'propositional contents' are objects of the so-called 'propositional attitudes'. This contrast between content and attitude yields a division of assertion into a sheer descriptive component and a sheer evaluative one and thereby invites relational analyses by virtue of this division. Yet according to Frege's Insight such division is rather misleading for our purpose to understand the assertoric force since it seems to suggest that the force is an additive ingredient to the content, as we usually consider what an attitudinal stance is. Hence, I will not adopt the distinction between propositional content and propositional attitude in this dissertation.[20]

At this point it should also be noted that even though the notion 'assertoric' was used in traditional Aristotelian logic to denote one of the three kinds of propositions that differ in modality (alongside 'problematic' and 'apodictic' propositions), Frege's notion of the assertoric force, however, is *sui generis*. It is hence neither any specific grammatical mood or mode of expression standing parallel to others like the declarative, assumptive, imperative, optative, permissive, etc. (I shall say more on this in §4.1.3.)

A final terminological issue I would like to mention is that when Frege uses the notion 'force' he intends to emphasize that when the content takes the form of assertion (i.e. becomes assertoric) it performs an act of applying a concept; one may view such a performance as an instance of manifesting the *power* of asserting or judging. This may lead to further questions of understanding the relation between 'power', sheer 'capacity' (i.e. potentiality), and their manifestations. This topic could become rather complicated.[21] I will not pick up the notion of power until §4.2.4, and a final remark of it will be postponed until §6.3.

After conditioning these terminologies, our task now is to give a satisfactory account of the assertoric force and its relation to the conceptual content. To begin with,

---

[20] Though not talking about attitudes, I do support a view of discursive commitment as well as the idea of commitment attribution (cf. Chapter 4). And I believe a successful treatment of the propositional attitude should agree with the treatment of discursive commitment outlined in my dissertation.

[21] This distinction is linked to the problem that notions like 'judgment' or 'assertion' actually reflect a fourfold ambiguity on this distinction: An assertion can either be understood as a type of repeatable act of judging, or an instance of this act-type, or a type of the 'product' of judging act, i.e. the asserted content, or an instance of this product-type. I will have more discussions in Chapter 2 about this complication.

in the next section I will first discuss the approach that interprets the assertoric force in psychological terms, which Frege considers to be severely mistaken.

1.3 Argument against Psychologism

One consequence of Frege's Insight is that the force obviously must not be at odds with the semantic truth of a content (since it rather constitutes the content's truth). It means that the force cannot be treated as some additive ingredient to the content that is conceptually inscrutable or non-evaluable in itself, such as certain psychological propensity or other non-cognitive attitudes[22]—Frege resolutely opposes reducing the assertoric force to any kind of psychological mechanism.

Consider Jack's example again: According to our discussion in §1.2, Jack's dismissing response to Sophie's attempt to give him reasons to believe in climate change was said to be self-defeating. However, it might be suggested that Jack's response needs not to be a total rejection of the possibility of meaning and communication, but he could nevertheless frustrate his interlocutor and dismiss her claim of fact by emphasizing the relativity of everyone's beliefs—Having recourse to psychological mechanism is an example, which contends that one's beliefs are influenced by non-cognitive conditions that do not constitute the content of those beliefs.

In this case, Jack disqualifies his interlocutor's assertion without abandoning the possibility of meaning in general. By appealing to the external conditions of Sophie's belief he is actually referring to a meta-level fact about Sophie's act of believing. Of course, we can disqualify an act of speech or belief—For example, we may dismiss one's utterance by pointing out that he is drunk—But for the very same reason such a dismissal will not be qualified as a discourse or conversation with him. As I will elaborate in Chapter 5 and 6, to disagree with someone we need to be engaged in a discursive explanatory act of reason-giving which must recognize at least part of the interlocutor's background beliefs as valid.

Any form of explanation is still a thought. Jack's approach of putting forward external qualifications of one's belief should nevertheless eventually be revolved as

[22] It means, moreover, that if the force is to be treated as having pragmatic effect (as I will be arguing for), then such pragmatic effect cannot be just some 'post-propositional' add-on, e.g. supplementary enrichment of the meaning of a proposition. See Chapter 2.

inside the global context of mutual intelligibility. Through this, we may see how Jack can do better: To achieve a good, meaningful conversation, one who dissents from the statement for climate change must bring forth evidence that contradicts that statement and in effect argues that climate change is not a fact. If one is going to have recourse to any further qualification or disqualification that contends "climate change is only *your* fact, but not a fact in *my* community or according to *my* standard", then either he is indeed appealing to an underlying fact which must be further elaborated, or he is not addressing the same issue and not really engaged in a conversation with his interlocutor at all—Although it is indeed sensible to point out that our cognitive acts are affected by pathological and psychological conditions, *reducing* the cognitive reasons to the pathological and psychological conditions, however, is putting the cart before the horse.

Frege is well aware of the primacy of cognitive reasons in discourse. He persistently emphasizes that we should avoid explaining thoughts and their laws in terms of the subject's psychic or dispositional states, desires, or other pathological conditions, theses he attributes to the so-called psychologism. In fact, he may count as one of the most prominent anti-psychologistic philosophers in history. (E.g. his critical review of Husserl's *Philosophie der Arithmetik* is sometimes believed to have turned Husserl away from psychologism.) For Frege, the performances of asserting or evaluating the thoughts cannot be explained by anything from 'outside' the Concept Script but only in terms of the thought's self-evaluation of its lawfulness.

In his book *The Last Word*, Thomas Nagel made a similar claim, which he summarizes as follows:

> The impulse to qualify is very difficult to suppress. The only way to resist the constant temptation to give the last word—even the unsayable last word—to the first person, singular or plural, is to see how first-order reasoning about the world inevitably dominates these ideas if we take them seriously. (Nagel 1997, 35)

Frege takes the Concept Script as a project to articulate laws of thought radically against any form of external qualifications. In rejecting psychologism, the Concept Script must stipulate the laws of thought in a way that it can also explain their applications. (As will be discussed in §2.1, this is directly linked to the problem of rule-following.) It means that if there are certain rules or norms that qualify a thought—including, e.g., the psychological ones—the instantiation of such rules or norms itself should also be articulated in thoughts, so that the whole process of qualification may take a form of

inference or reasoning, which is essentially first-order and licensed in conformity with the Concept Script.[23]

1.4 Deflationism and Its Problem

1.4.1 Deflationism, Regress, and Truth

The argument against psychologism, as we saw in the last section, is an argument against external qualifications of the performances of asserting or evaluating thoughts as according to their norms. It infers that even though we can explain or evaluate a thought in a language by having recourse to a meta-linguistic explanation, i.e. by virtue of 'semantic ascent', the ascent does not help us close a chain of explanation. In other words, the semantics should not be *inflationary* with regard to explanations of the performances as instantiations of norms.

One famous example of inflationism is the modus ponens reasoning characterized by Lewis Carroll in his *What the Tortoise Said to Achilles*: An interpretation of the modus ponens is supposed to take the form that if I assert *q* as a premise and conclude that *p*, then it implies a hypothetical judgment "If *Q* then *P*"—insofar that my act of inference is a rule-following act. But this interpretation gets us nothing but a third extra assertion that needs to be further explained if we do not want to beg the question. Generally, as Timothy Williamson summaries:

> We can distinguish two levels of reflection, the logical and the metalogical […] The logical level is not purely mechanical. When the reasoning is complex, one needs skill to select from the many permissible applications of the rules one sequence that leads to an answer to the question […] One starts only at the metalogical level of reflection to think about the semantics of the logical connectives and other expressions one employed at the logical level. [...] It must be possible to think logically without thinking metalogically, for otherwise by the same principle thinking metalogically would involve thinking metametalogically, and so ad infinitum. (Williamson 2007, 40f.)

Rejecting inflationism does not mean that there could be no semantic ascent whatsoever, such as reflections or critical examinations of our interpretative uses of contents and their meanings. But since Frege's idea of the Concept Script is to propose a singular

[23] In Chapter 6 I will give a more detailed discussion of reason-giving acts with qualifications, which, I argue, are inherently pedagogical.

language of pure thinking that enjoys universal validity of meaningful applications, the ascent cannot take place at the level of the interpretative understanding of inferential rules or the possibility of a universal semantics of the language of thought *as such*.[24] If we always have to have recourse to external qualifications to ground our conceptual use in order to speak meaningfully, we would not be able to speak of anything meaningful at all; and just like the Achilles who can never catch the tortoise, a meaningful performance of our thinking could never get started.

This idea coincides with Frege's Insight that the assertoric force is neither additive nor assimilable to the content. In his later years, Frege states in a fragment dated around 1915 titled "My Basic Logical Insights":

> Making a judgment does not alter the thought that is recognized to be true. […] The word 'true' is not an adjective in the ordinary sense. […] So the sense of the word 'true' is such that it does not make any essential contribution to the thought. […] 'True' only makes an abortive attempt to indicate the essence of logic, since what logic is really concerned with is not contained in the word 'true' at all but in the assertoric force with which a sentence is uttered. […] (Frege 1979, 251f.)

Frege emphasizes that the word 'true' that marks the assertoric force does not really have any linguistic content—There are, of course, specific contexts in which we talk about truths or facts in a substantial manner. When one says: "the climate change is a fact", the word 'fact' may indeed invoke some specifically scientific properties such as observational evidence; or when I claim that "Goldbach's Conjecture is true", I thereby imply that there exists a mathematical proof of Goldbach's Conjecture—But in Fregean contexts, the truth expresses no semantic content when it is prefixed to assertions like 'It is true that *p*', which is the same as the assertion *p*. In contrast to inflationism, we may call such approach a *deflationary* one regarding the semantics of truth.

With deflationism, it is sometimes debated that the concept of truth is simply redundant. But before we could take on such a debate, we need to make it clear that it is—according to Frege's Insight—not primarily an issue about what kind of semantic content the concept 'truth' may or may not have, but rather an issue about what kind of role does the concept of truth constitutively play in order for us to understand any

[24] Cf. Hilary Putnam: "There is no 'metalanguage' in Frege, in which we could say that the laws of logic are 'logically true'; one can only assert them in one language, *the* language". (Putnam 2000, 220).

content's semantics in general.[25] This point has been explicitly indicated in Frege's passage quoted above: the expression of truth signifies nothing but the assertoric force in utterance. If Frege is correct, then what we really need to comprehend is the assertoric force, while the meanings of notions such as 'truth' are rather auxiliary and dependent upon our comprehension of the force. (Later in this dissertation I will argue that for someone to assert *p* it is essentially equivalent for her to assert '*p*', or '*It is true that p*', or '*I take p to be true*', or '*I think p*', or '*I believe p*' etc.[26])

Here, it suffices to see that according to Frege's Insight taking a content to be true is not predicating a property of that content. And since the expression of truth is inherently indicated by the force of discursive utterance, there is no essential difference between a content's of being *taken to be true* on the one hand and the content's *being true* on the other. In fact, the primacy of the assertoric and discursive articulation of conceptual contents should not only *not* be taken over by the psychological mechanisms that qualify our discursive acts from outside, but also *not* be ceded to the ontological facts of certain self-contained entities as truth-makers that are supposed to be independent of our discursive practices.

But it is not to deny that there are some serious ontological concerns if we are to equate something's objectively being true with its 'merely subjectively' being taken to be true by someone—Our practice of knowledge and inference does not stop at just correctly understanding what someone *literally* says but inherently concerns evaluating them as expressing a fact or not.[27] For example, understanding what literally is said by either the expressions "The current French king is worried." or the expression

---

[25] It might be helpful to distinguish the word 'true' on the one hand, and the notion of truth on the other, and take only the latter as belonging to the investigation of the force and hence the essence of logic. See §1.4.4. Cf. also Dorit Bar-On and Keith Simmons: "We think that Frege teaches us a very important lesson. Deflationism about the word 'true' is one thing, deflationism about the concept of truth quite another." (Bar-On and Simmons 2007, 76).

[26] It will be explained §4.3, above all, in terms of the discursive agency and its commitment at play in the performance of the assertion, which exemplifies what I call the 'performative immediacy', namely, the idea that in the performance of asserting something to be true there is no difference between the description of the act of asserting or believing that content and the actual act of asserting or believing.

[27] See my discussion in §§5.4–5.5 on the two-dimensional semantics and cross-contextual evaluations and disagreements. On the tension between semantic truth and ontological facts cf. also Kit Fine (2007, 44), where he distinguishes between "semantic facts" from "semantic truth".

"Hesperus is the brightest object in the sky before sunrise." does not yet yield us any truth nor knowledge, since it does not correspond to any fact.

For Frege, the question of meaning is different from the question of understanding the literal or lexical meaning of an expression. He understands his Concept Script as exclusively a logical project of accounting for the laws of what it means for the content to be true—He argues that while all our sciences should take the form of being true, when it comes to logic, it not only bears this form, but also has this form as its object of study. Frege begins his *Logical Investigations* by pointing out this idea:

> All sciences have truth as their goal; but logic is also concerned with it in a quite different way: logic has much the same relation to truth as physics has to weight or heat. To discover truths is the task of all sciences; it falls to logic to discern the laws of truth. (Frege 1918–19, 58)

For Frege, the concept of truth is nonetheless indispensable because it is the object of study of his Concept Script, which investigates the "laws of truth", i.e., the "general features […] of what is" (ibd.). But how should we understand such a general feature of being true if it is not a property? In *Begriffsschrift*, Frege declares:

> You may if you like distinguish subject and predicate even here; but the subject contains the whole content, and the only purpose of the predicate is to put forward (*hinstellen*) the content in the form of a judgment. *Such a language would have a single predicate for all judgments, viz. "is a fact"*. (Frege 1879, §3)

According to Frege, if we call the whole conceptual content the subject, then the assertoric force denoted by the judgment-stroke—astonishingly—can nonetheless be seen as a predicate. But isn't it also Frege's own insight that the force cannot be assimilated to predicates? One possible response would be that the force is a special, i.e., meta-linguistic predicate, but it seems to invite an inflationary conception of the force, e.g. as a function of interpreting a function, or a predicate explaining predication, which induces regress.

In sum, as a supposedly universally intelligible language, the Concept Script is conceived to be able to eliminate all kinds misunderstandings, false translations, vagueness, and homonymy in our natural languages. Moreover, Frege contends that a content's universal intelligibility means nothing different from its being true, i.e. being a fact—For a content to be universally intelligible, therefore, is the sufficient and

necessary condition for it to be true—By rejecting inflationism and psychologism, Frege makes it clear that we need no extrinsic ontological condition or truth-maker to qualify a conceptual content and render it true; quite the contrary, the Concept Script must be *self-explaining* in terms of its truth or universal intelligibility. It reflects Frege's Insight that the assertoric force, and hence the concept of truth, is not assimilable to a predicate, which, in turn, explains why it is so hard to find any expression in ordinary language that regularly corresponds to the judgment-stroke in the Concept Script.

Nonetheless, we can still see an apparent difficulty here—as the example of "Hesperus is the brightest object in the sky before sunrise" indicated—between the condition of a content's 'literal' intelligibility on the one hand and the 'factual truth' expressed by the content on the other. The question is: Can such a statement be of any cognitive, conceptual value at all? If yes, how?

To put it in another way: To follow Frege's Original Insight, we can take the notion of *truth* to be equivalent to the notion of *discursive appropriateness* (In fact, I will adopt this equivalence in this dissertation) and thus for a content to be true it means nothing other than the discursive appropriateness of uttering the content at its context, namely, nothing other than the uttered content's being a content proper. If we adopt this view, it means that the notion of truth in this general, discursive, and assertoric *force-indicating sense* must be distinguished from the notion of truth *as a kind of truth-value*, which is rather specific and technical. Both the statements "P has the truth-value 'True'" and "P has the truth-value 'False'", in so far that they are discursively appropriate statements, are therefore true. That is exactly why, for Frege, the force-indicating assertion-sign cannot be further nested by any other signs, because it always already indicates the *scope* of the entire sentence. In other words, only with the assertion-sign can we say that the sentence is complete and express a unified thought.

With the differentiation between the two notions of truth, however, there is indeed a tension here: Since a thought must be subject to evaluations, which could be either affirmed or denied as being true. What kind of thought is it that we are denying, if it is, on the one hand, taken to be a false statement (having the truth-value 'False'),

while on the other hand, still considered as discursively appropriate? This leads to what I call 'Frege's Puzzle', which I will discuss in more detail in §1.4.3.[28]

Frege himself has obviously struggled to elaborate the idea of truth, and his invention of the judgment-stroke in his Concept Script—regardless of his emphases of its value—has caused confusion and met with much criticism from mathematicians to philosophers. We will take a closer look at this confusion in §§1.4.3–1.4.4, but before that, I would like to briefly sketch Frege's logical conception of the laws of thought. To do so, let us consider *§24* in his *Begriffsschrift* as an epitome of his idea, where he first introduced the idea of a theory of sequence (*Reihenlehre*), which is viewed as an idea of second-order predication and later became the basis of his theory of number.

1.4.2 The *§24* in *Begriffsschrift*

In *§23* of his *Begriffsschrift* (which prefaces the book's Part III: 'Some Topics from a General Theory of Sequences'), Frege claims that the Concept Script must be able to show …

> […] how pure thinking irrespective of any content given by the senses or even by an intuition a priori, can […] bring forth judgment that at first sight appear to be possible only on the basis of some intuition. This can be compared with the condensation, through which it is possible to transform the air that to a child's consciousness appears as nothing into invisible fluid that forms drops. (Frege 1879, §23)

This passage expresses two ideas: *First*, the Concept Script is a language of pure thinking, which is responsible for demarcating whether a content given by the senses—e.g. the content that "the current French king is worried"—brings forth judgment or not. *Second*, such demarcation can be viewed as a 'transformation' from a 'mere intuition' to a judgment—comparable to a pedagogical movement, i.e., a movement from a thinker's ignorance to clarity of the truth-value of a content.[29]

In *§24*, Frege contends that although such demarcation is carried out in terms of pure thinking, it is not a judgment, but an *explanation* (*Erklärung*) of *content-identity*,

[28] In short: Although Frege correctly distinguished between *a)* truth that indicates the assertoric force and *b)* truth as a truth-value, he nonetheless fell short of explaining why they are sharing a same assertion-sign in his *Begriffsschrift*.

[29] It is subject to debate what Frege means by 'intuition' (especially in relation to Kantian philosophy). For simplicity purposes, I take it here to be simply meaning non-conceptual sensory content e.g. qualia, etc.

which indicates how various terms or entities are considered as conceptually equal—Such an explanation can be written as follows:

$$(\textit{Formula 69}) \quad \forall x\,(F(x) \rightarrow \forall y\,(f(x, y) \rightarrow F(y))) \equiv \mathrm{I}\,(F, f)$$

which says that the property *F* is inherited in the *f*-sequence.[30] Frege explains that this formula …

> […] differs from the judgments hitherto considered in that it contains signs that have not been explained before; it itself gives the explanation. It does not say: "the right side of the equation has the same content as the left"; but: "it *should* have the same content". Hence this proposition is not a judgment […] Such explanations only aim to bring about an external simplification by stipulating an abbreviation. […] Although originally (69) is not a judgment, it is immediately transformed (*verwandelt*) into one; for, once the meaning of the new sign is stipulated (*festgesetzt*), it must remain valid, and therefore formula (69) also holds as a judgment, but as an analytic one […] This *dual character* (*Doppelseitigkeit*) of the formula is indicated by the use of a double judgment stroke. (Frege 1879, §24, my emphasis)[31]

Again, two ideas are expressed §24: *The first idea* is that which we are nowadays familiar with—namely the introducing of the *universal quantification*, which is remarkably significant to Frege's overall philosophical undertaking:[32] Frege's approach of universal quantification is based on his interest in looking for the logical foundation for natural numbers in terms of a general theory of sequence. He introduced the quantifier to describe the domain of the arguments that can be asserted by a function—Yet the domain of the arguments is unconstrained since he only introduced one quantifier: the universal quantifier.[33] Therefore, while truth is characterized as a 'special predicate', or 'quasi-property' that every judgment shares, the demarcation of a sensory

---

[30] See Frege (1879), §24. Here the Formula 69 has been transcribed using the modern notations.

[31] The double judgment-stroke is the sign '||—', which is sometimes taken as a definition-stroke.

[32] Dummett once lauded that Frege's discovery of quantification is the "deepest single technical advance ever made in logic." See Dummett (1981, xxxiii).

[33] Strictly speaking, the quantification of *x* may be considered bound by the function F since F(*x*) appeared in the antecedent of the conditional; but the quantification of *y* is not constrained by F (since F(*y*) only appeared in the consequent). The variable *y* is constrained by F only when the sequence *f* inherits F, but this is only circularly shown in the formula by a mutual definition between the quantified scope of the procedure of extending the sequence and the scope of inheritance.

content as a thought is carried out by virtue of a judgment that classifies the content as an instance of object that is conceptually equal to an open, infinite sequence of objects.

According to such an approach, a content is true iff every instance of its interpretation renders the self-sameness of the content. For example, since 'apostle' means either Peter, or John, or Matthew …, we may write them schematically as a_1, a_2, … a_{12}. Yet while 'apostle' is a concept, every instance of interpreting 'apostle' is considered to be an act of representing an object in a quantitatively unconstrained sequence, e.g. a_i. It is then compared to a_1 through a_{12}—if a_i is not identical to any of the 12 objects, then it is a false interpretation.

But we need to understand why and how do I have to subject my instance of interpretation of 'apostle' to the extended sequence of counting *that* kind of object in the first place, namely that the instantiation of which is schematized as the '*a*'s. Frege's answer is that my interpretation of 'apostle' necessarily falls within the sequence of '*a*'s because the sequence is an unconstrained domain of objects, so that the *a* can be seen simply as a variable. However, an unconstrained quantification will also mean that an instance of interpretation is not distinguishable from other instances since there is no further category available to classify them—other than being instances of *the* kind of objects (i.e. objects of purely numerical differences) *as such*. In fact, since an instance of interpreting a concept can be understood no more than the purely arithmetic act of counting as such, a pure act of counting does not pick out an object in reality (i.e. object that is not purely arithmetic) and hence it does not deliver any semantic truth of a concept.

Moreover, an object cannot be assimilated to the instantiation of one concept and that concept only.[34] In reality, the word 'apostles' means more than its number '12'. But since Frege's calculus only applies to an unspecified domain of countable objects the concept of which is supposedly univocal—in the manner that the number of apostles and the number 12 are notoriously symmetrically substitutional in all contexts in the Concept Script—it is left unexplained how propositions such as Formula 69 could be

[34] As Brandom pointed out: "'thing' and 'object' are pseudosortals […] used when for some reason […] one does not want to specify the relevant sortal explicitly." (Brandom 1994, 438) If I put a book on the desk and ask "How many?" the answer could be 1, or 300, or 30000 concerning whether that which I meant to count is the book, or its pages, or its words; but according to universal quantification, my answer would by no means be constrained but that I simply count "five instances" without specifying what those five instances are of. But for this reason, my answer is obviously unintelligible.

applied to any domain of objects other than the purely arithmetic one.[35]

Nevertheless, I would like to suggest that the *§24* in *Begriffsschrift* also expresses *a second idea*, which is more easily neglected, namely that the Formula 69 has a '*dual character*' (*Doppelseitigkeit*): It expresses *in a sense not* a judgment but a prescriptive content (which is sometimes interpreted as a 'definitional commitment') yet *in another sense* indeed a judgment. Frege's characterization here is very interesting. It indicates that what he is attempting to achieve in *§24* is not simply to propose—as generally assumed—a second-order predicate logic. Just like that truth is deemed only a 'quasi-property', for Frege the content-identity expressed by Formula 69 is only an analytic judgment iterating in a sequence what was originally recognized as identical in content, while he refuses to regard the recognition itself as a judgment but only as something prescribing heredity.[36]

We can find Frege reaffirming the prescriptive, i.e., normative character of pure thinking in his *Grundgesetze*, where he maintains that the laws of thought are those that may prescribe how one *ought* to think:

> The double sense (*Doppelsinn*) of the word 'law' is fatal here. In one sense it says what is, in the other it prescribes what ought to be. Only in the latter sense can the logical laws be called 'laws of thought' (*Denkgesetze*) insofar as they stipulate (*festsetzen*) how one ought to think."[37]

Interestingly, here again Frege claims that there is a '*dual character*' for the laws of logic: In a descriptive sense, the laws of logic can be seen as the 'laws of truth' that describe the general feature of what is; while only in a prescriptive sense can the laws of logic be viewed as 'laws of thought'—The two senses are interdependent, and both of

---

[35] Cf. e.g. Rödl (2012, 49), and Fine (2002) chap. 1. Even in mathematics, as we know, it is exactly the idea of a singular universal domain of objects that gave birth to the Russell's Paradox.

[36] We can see that this is consistent with Frege's idea put in *§23* that the demarcation of meaning can be compared to the pedagogical movement. In Chapter 6, I shall say a bit more about the significance of the idea of pedagogy in the dynamics of belief revision, according to which certain norm is viewed as in a sense non-constitutive and yet in a sense constitutive of a discursive act.

[37] "Der Doppelsinn des Worts 'Gesetz' ist hier verhängnissvoll. In dem einen Sinne besagt es, was ist, in dem andern schreibt es vor, was sein soll. Nur in diesem Sinne können die logischen Gesetze Denkgesetze genannt werden, indem sie festsetzen, wie gedacht werden soll." (Frege 1893, S. XV)

them are needed if we are to understand the significance of the Concept Script:

In one sense, the Concept Script merely describes the constitution of a sentence in it. To describe a sentence as conforming to the laws stipulated by the Concept Script is to pass a judgment on its truth. But such judgment would be redundant if the sentence is tantamount to the laws themselves, hence we need a normative explanation how the sentence is bound by the laws while not being tantamount to them. Therefore, the Concept Script, in another sense, also prescribes those which ought to be constitutive of a sentence insofar that the sentence could be written in the Concept Script, that is, insofar that we are to explain the sentence as instancing a thought in the first place (rather than, say, randomly scribbling on a paper, making a noise, or involuntarily making an arm gesture because of a muscle twitch, etc.). Such an explanation, according to Frege, is not a judgment, but only a normative explanation of heredity; yet on the other hand, the normative explanation can be conceptually understood as nothing but judgment that demarcates a sentence as thought-conferring according to the laws stipulated by the Concept Script.[38]

However, these two senses of the laws seem to be at odds with each other (if not viciously circular): On the one hand, for Frege, the laws for demarcating thoughts and the laws for their truths are the same, and hence only lawful thoughts can be demarcated as thoughts—In mathematics, if I violate certain set of axioms, it does not necessarily follow that I am thereby incapable to confer any meaning or thought; I might, for example, be working on another axiomatic system—But if I violate a law of thought, according to Frege, I would not be conferring any thought at all. Yet on the other hand, since a demarcated sentence in the Concept Script may or may *not* have the truth-value 'True', it would be problematic to tie up the question of demarcation with the question of lawfulness of thought. Although it is stipulated by the Concept Script that an instance of expression conferring a thought must conform to the laws of thought, it must be possible for the expression to fail to conform to some of the laws of thought—while being yet still intelligible and thus demarcated from sheer nonsense.

To solve this tension, Frege introduced the famous distinction between an expression's sense (*Sinn*) from its reference (*Bedeutung*). The significance of this

[38] In line with such a 'dual character', recently some philosophers have argued that we should regard the laws of logic neither as being sheer constitutive (i.e. descriptive) nor as being sheer normative but as being 'constitutive-normative'. Cf. e.g. Leech (2015), and MacFarlane (2002).

distinction is twofold: First, it squares with the idea that there must be more than one way (namely, more than one sense) to pick out a referent; Second, it allows space for non-omniscient and fallible thinkers such as us human beings to be qualified for having thoughts.

But it is unclear how the sense–reference distinction could be compatible with the basic insight in his *Begriffsschrift*. At first glance, the solution might seem to be straightforward: Recalling the distinction between the horizontal content-stroke and the vertical judgment-stroke, we might say that a content that has a sense can be prefixed by the content-stroke—as a content that is *truth-apt* and *judgable*, and hence demarcated from other intrinsically non-sensical or non-judgable contents—such truth-apt content may or may not be asserted as true; and when it is asserted as true, it will be further prefixed by the judgment-stroke and therefore has a reference. However, such a construal conflicts with Frege's Insight that a thought is demarcated by, and only by, the assertoric force (i.e. when it is asserted). Therefore, it raises the question how it is possible to demarcate a thought if it is not asserted. I will argue that it is indeed a serious puzzle for Frege's project.

1.4.3 Frege's Puzzle

We can summarize Frege's Puzzle as follows: A thought is demarcated *as thought* iff it is universally intelligible, and it is stipulated by the Concept Script that its universal intelligibility is sponsored by the assertoric force, which is indicated by the assertion-sign. However, it seems that the assertion sign plays simultaneously two roles here: On the one hand, it assigns a sentence a truth-value; yet on the other hand, it also asserts the sentence, namely, the sentence is said to be true. In other words, it might seem that a thought's *universal intelligibility* (as a judged fact) and its local *semantic stability* (as a merely evaluable thought) in conversations seems to be not really overlapping,[39] because a semantically stable, truth-evaluable content is considered nonetheless fallible.

Frege's distinction between the judgment-stroke and the content-stroke in his *Begriffsschrift* does not seem capable of accounting for this distinction. It would be

[39] We might say that the content's universal intelligibility stipulated by the Frege's *Begriffsschrift* exemplifies 'semantic superstability' (content with semantic superstability has a literal meaning that always picks out the same object under every interpretation) as contrasting with a content's customary conversational semantic stability at given contexts. More discussions on this see Chapter 2.

wrong to think that the content-stroke may serve as that which turns a non-judgable content into a judgable one—The content-stroke does not serve for demarcating contents, since the whole purpose of the *Begriffsschrift* is to emphasize that to demarcate a content as universally intelligible is nothing but to pass a judgment on its truth.

As we have briefly indicated earlier, a careful treatment has been adopted by Frege, with his Original Insight, to distinguish between the force-indicating truth understood as universal inferential intelligibility in discourse and the truth understood as merely a truth-value. In his *Grundgesetze*-period, he tried to incorporate the truth-value reading and the inferential reading of truth into a uniformed logical picture, where he distinguished between concepts and objects—A concept *simpliciter*, he argued, is a function that remains 'unsaturated'; and it turns 'saturated' if it is interpreted by an argument that is an object, which is always saturated; and eventually, some saturated concepts can be judged to be either true of false by a characteristic function, by virtue that they refer to either of the two special kinds of objects: the truth-value True and the truth-value False—In this rather technical picture, the content-stroke is supposed to play a unifying role of turning a saturated concept, e.g. a singular term, into a judgable sentence. However, such a functional treatment of concept does not tally with, but rather has largely betrayed, the notion of conceptual content in *Begriffsschrift*, since unlike concepts, which are not self-subsistent, the content-stroke always prefixes a complete sentence rather than an uninterpreted function—As Dummett puts it:

> The content-stroke could be meaningfully prefixed only to an expression for a judgeable content […] Thus in *Begriffsschrift* the content-stroke does not serve the purpose of turning an expression for an unjudgeable content into one for a judgeable content, whereas in *Grundgesetze* it in effect turns any singular term into a sentence. In *Begriffsschrift* Frege says that the content-stroke serves both to combine the symbols following it into a whole, and to relate any sign to the whole formed by the symbols following the stroke. But these two operations are quite redundant: whatever can meaningfully follow the content-stroke must already constitute a unitary whole, and any symbol that can be joined on to the content-stroke must have already been explained as capable of being related to such a whole. In any case, the content-stroke is superfluous […] (Dummett 1981, 315)

Abandoning the content-stroke does not give us a solution to Frege's Puzzle, however. Bertrand Russell, for example, abandoned the content-stroke and used the assertion-sign

as a whole (i.e. no longer distinguishing the horizontal from the vertical). But he was nevertheless caught by the difficulty in accounting for the relation between the demarcated unity of thought on the one hand and its components that do not have such unity but remain nonetheless thinkable on the other. He says:

> A proposition, in fact, is essentially a unity, and when analysis has destroyed the unity, no enumeration of constituents will restore the proposition. The verb, when used as a verb, embodies the unity of the proposition, and is thus distinguishable from the verb considered as a term, though I do not know how to give a clear account of the precise nature of the distinction. (Russell 1903, §54)

Since the force that demarcates the unity of a thought is indicated by an act, Russell suggests that we can regard the unity as being embodied in a verb; yet such a force-indicating verb must be differentiated from a regular verb as a term. The difficulty is that it remains unclear how such distinction could be drawn—Just like Frege's Insight that the force cannot be assimilable to predicate, which is a merely negative statement, the demand to make such distinction is not yet a solution.[40]

According to the *Begriffsschrift*, the meaning of any verbal content or component of a thought is derivative of the meaning of a demarcated sentential unity. Also according to the *Begriffsschrift*, sentences "Phosphorus is Phosphorus" and "Phosphorus is Venus" express the same thought since they refer to the same fact, and their difference belongs only to the 'wording' or 'coloring'. However, for a non-omniscient thinker, the two sentences seem to have an essential cognitive difference, which is constitutive of our assertion on their meanings and hence constitutive of their demarcation, rather than a sheer rhetoric difference that is merely post-propositional, additive, and not constitutive of their meaning-demarcation. To solve the problem, it seems that we need to maintain that the force is meaning-constitutive but rejects that the two sentences above have the same meaning. Yet it is unclear how Frege's assertoric force could account for the cognitive significance of the differences among the senses of non-omniscient thinkers. We saw that he attempts to explain this in terms of the prescriptive character of the heredity of contents in *§24*, but he eventually ends up

[40] To balance between semantic stability and fallibility Russell later proposed the multi-relational theory that parameterizes the instances of applying concepts, according to which 'I think *p*' would be described as *p* having a property of being thought (being interpreted) by me, or equally, me standing in a relation to *p* that exemplifies a two-place function of thinking. But such an approach obviously went off course from Frege's Insight and seems inflationary.

defining assertions with analytic truths (e.g. by employing a calculus based on unconstrained quantification of content-identical instances). In fact, during Frege's lifetime, especially after the devastation brought by Russell's Paradox that shook Frege's overall logical project to its very foundation, it became ever doubtful among philosophers if the assertoric force could be incorporated into logic at all.

Wittgenstein, for example, argues in his *Tractatus*:

> Frege's judgment-stroke is logically altogether meaningless: in Frege (and in Russell) it only indicates that these authors hold the sentences marked with this sign to be true. Thus [the judgment-stroke] is no more a component part of a sentence than is, for instance, the sentence's number. It is impossible for a sentence to state (*aussagen*) by itself that it is true.[41]

Frege would probably agree with the last sentence of Wittgenstein's criticism above—For Frege, it is exactly because a 'mere' statement (*Aussage*) is not semantically autonomous that we need a force to assert it—Their main disagreement is that while Wittgenstein was overall pessimistic about incorporating the assertoric force into a project like the *Begriffsschrift*, Frege, even in his later years, resolutely believed that the force not only is needed but also is even the essence of logic. As an example, he ends the manuscript "My Basic Logical Insights" with the following passage:

> 'True' only makes an abortive attempt to indicate the essence of logic […] Now the thing that indicates most clearly the essence of logic is the assertoric force with which a sentence is uttered. But no word, or part of a sentence, corresponds to this; the same series of words may be uttered with assertoric force at one time and not at another. (Frege 1979, 252)

Yet Frege seems to be suggesting something hopelessly aporetic here: On the one hand, he suggests that the essence of logic is the assertoric force instead of 'truth'—Namely, the assertoric force is more crucial for us to understand logic, i.e. the laws of thought; while the word 'true' only makes an 'abortive attempt' to reach such understanding—

[41] "Freges 'Urteilsstrich' ist logisch ganz bedeutungslos; er zeigt bei Frege (und Russell) nur an, daß diese Autoren die so bezeichneten Sätze für wahr halten. [Der Urteilsstrich] gehört daher ebensowenig zum Satzgefüge, wir etwa die Nummer des Satzes. Ein Satz kann unmöglich von sich selbst aussagen, daß er wahr ist." (Wittgenstein 1921, §4.442).

On the other hand, our ordinary languages are so imperfect that we do need help from the notion of truth, and our need of logic rises exactly from this imperfection.[42]

The 'abortive' character of the word 'true' indicates that any expression in our ordinary language conceived to convey the assertoric force would be eventually futile:

> We express acknowledgement of truth in the form of assertoric sentence. We do not need the word 'true' for this. And even when we do use it the properly assertoric force does not lie in it, but in the assertoric sentence-form; and where this form loses its assertoric force the word 'true' cannot put it back again. (Frege 1918–19, 63)

Frege's example is that, just as a stage thunder is only a mock thunder and a stage fight only a mock fight, a stage assertion is only a mock assertion—In general, we can say that a thought in fiction is a mock thought (*Scheingedanke*)—One has to speak *sincerely* or *wholeheartedly* in order to apply the assertoric force. But how do we know whether a sentence is really an assertion, if there can be no word or indication whatsoever in the sentence that may guarantee that it is seriously asserted? There seems to be no definitive answer to this.

1.4.4 The Problem of Fiction and Distanced Analyzability

The problem is directly derived from the tension between Frege's two Insights: on one hand the force is meaning-constitutive, yet on the other hand it makes no contribution to the content. A common but problematic approach is to neglect the first insight and declare that the force is either redundant or conceptually inscrutable, which infers that a content could be seen as semantically autonomous and self-contained, i.e. as an objective and 'force-free' entity with its truth-condition nonetheless intact regardless of its different uses in different contexts, while the force would be a merely 'entertaining' attitude that somehow added or attached to that content. In fact, Frege himself appears to have also adopted such an approach—most significantly in his late writings—which is sometimes labeled as the 'Frege Point', as Geach once summarizes:

---

[42] Cf.: "How is it then that this word 'true', though it seems devoid of content, cannot be dispensed with? […] That we cannot do so is due to the imperfection of language. If our language were logically more perfect, we would perhaps have no further need of logic, or we might read it off from the language." (Ibd.)

> A thought may have just the same content whether you assent to its truth or not; a proposition may occur in discourse now asserted, now unasserted, and yet be recognizably the same proposition. (Geach 1965, 449)

This seems to square with Frege's own statement that "the same series of words may be uttered with assertoric force at one time and not at another" (1979, 252). However, in later chapters I shall argue that the Frege Point is very misleading and contradicts Frege's Original Insight, and we can by no means consider a 'force-free' content as semantically autonomous. Here, I would like to briefly indicate that as long as we maintain Frege's first insight that the force is always meaning-constitutive, it should be clear that the reply to the problem of fiction consists in maintaining that a fictional discourse actually requires *more*, *not less*, qualifications for its truth-condition. (Philosophers usually calls the qualification for unasserting a thought '*cancellation*', but in this dissertation I will use the notion *distancing* to refer to such qualification.) Anyways, we should see that for an unasserted content its assertoric force is canceled or distanced *not* in the sense that the content is hence 'force-free'; rather, the content receives further qualification *on top of* the asserted content.[43]

More detailed discussion on the force and the distancing must be postponed to later chapters. Here let me just recapitulate the problem inflicting on the project of accounting for the assertoric force:

The assertoric force was proposed to give an answer to the *demarcation question*. A Fregean answer is hence that we demarcate a content iff it is demarcated by the assertoric force—According to Frege's Original Insight, it is the assertoric force alone that demarcates unity of a content; and only a complete sentence rendered by an act of assertion bears the force, which enables its content to be demarcated as meaningful—In other words, the force makes sure that the assertion enjoys explanatory primacy over its constituents in terms of this very unity.

However, we have also seen that there is a second question, i.e., there seems to be other ways to demarcate a thought than asserting it. A content should remain thinkable even if it is 'unasserted' (e.g. when it is denied, or fictively uttered, or

[43] Perhaps the derivative nature of the unasserted content can also be applied to our understanding of Frege's notation in his *Begriffsschrift*: The content-stroke, which expresses the self-sameness of a content, does not need to be viewed as redundant. Instead we can regard it as a sign derived from the assertion-sign indicating the cancellation of the force.

embedded in another assertion, etc.).[44] Hence if a content can only be demarcated by the assertoric force, we need to explain how we could understand an unasserted content and differentiate it from asserted ones—I would like to call it the *question of distanced analyzability*, which asks how it is possible for us to distance ourselves from an assertion without yet giving up the condition of discursive intelligibility of that thought. In short, we must be able to distinguish between three kinds of things: *a)* full-fledged assertions, *b)* unasserted contents, and *c)* intrinsically unintelligible contents.

In this dissertation, we need to understand how the assertoric force is meaning-constitutive, with regard not only to the demarcation question but also the question of distanced analyzability.

1.5 Assertion and Discourse

In §1.4 we discussed deflationism and its related difficulties. The deflationism of truth coheres with Frege's Insight to the extent that the assertoric force cannot be reduced to a predicate or a property; but it generally leads to the conclusion that the force is either redundant or not meaning-constitutive. We need to find a middle way that maintains Frege's Insight without embracing such a conclusion—We need to see that the assertoric force nonetheless plays a meaning-constitutive role *in discourse*.

In a discourse, context and background information become important for understanding meanings. Take fictional expressions for example, we differentiate whether an actor on stage is reciting a line in a play or talking on behalf of himself based solely on context, while our understanding of the context is based on further knowledge and assertions in our cognitive, discursive history. Hence the whole discourse on meanings can be viewed as inferences from multiple assertions—As Dummett comments: "The traditional answer to the question what is the subject-matter of logic is, however, that it is, not truth, but inference, or, more properly, the relation of logical consequence. […] and it is, surely, the correct view." (Dummett 1981, 432f.)

[44] I have not carefully distinguished between these cases of unasserted content—In Fregean contexts things could become slightly complicated, since for Frege, while a falsehood has a truth-value False, an empty name or a fictional sentence lacks any truth-value. I will not get into the details of this controversy in Fregean context. But I will touch on this issue in §5.5 by distinguishing between context-revision and context-shift (or equally, between the genuine and the faultless disagreement). Cf. also my discussion on 'polemic negation' in Appendix 1.

From such a discursive (or, broadly inferentialist) view of Frege's project, we can better understand why for Frege the essence of logic does not primarily concern what is true but the assertoric force, since only the force is inherently discursive.[45] In Chapter 2 and 3, I shall give further examinations about the force–content relation, and suggest that we must treat the assertoric force as inherently pragmatic. I will propose my speech-act-theoretic approach to the assertion in Chapter 4, where I argue that to assert is to make a performative commitment to the truth (i.e. discursive appropriateness) of a proposition at the context of discourse. The utterer's understanding of the proposition's truth, therefore, refers to the conformity to the epistemic standards and norms implicit in the background knowledge of the context. After discussing the difficulties and challenges to the speech-act-theoretic approach in Chapter 5, I will conclude in Chapter 6 that we understand how the assertoric force could be meaning-constitutive only through a 'dynamic' view of our discourse.

[45] As we know, the assertion-sign has been indeed used in logic as a turnstile for consequence. E.g. if one infer C from a set of background information Γ and premises A and B, we can use the assertion-sign as indication of the conclusion of the inference: $\Gamma, A, B \vdash C$. A more interesting reading of the assertion-sign, therefore, would be to think of it as belonging to the symbols of inference (e.g. together with other symbols for reasoning introduced by Frege in the *Grundgesetze §15*).

CHAPTER 2

The Essential Pragmatics of Assertion

This chapter argues that the speaker's pragmatic condition is constitutive of the truth-conditional content of the speaker's utterance. It therefore concerns issues about the semantic–pragmatic relation.

In Chapter 1, I have introduced Frege's idea of the Concept Script and his Original Insight into the assertoric force as an expected solution to the question of demarcation. However, we have also seen that there is another question, i.e. the question of distanced analyzability that demands us to account for demarcated yet unasserted thoughts. I showed that Frege was also puzzled by this demand. In what follows, I will further argue that the puzzle eventually led him to adopt a dualistic account of force–content relation, which consequently makes the rule-following problem inexplicable. After discussing Frege, I will explore several influential approaches to the problem of pragmatics. At the end of the chapter, I will reach a conclusion supporting J. L. Austin's speech-act based, performative theory to account for the essential pragmatic condition for a speaker to make an assertion

2.1 Frege's Dualism

2.1.1 Dualism and the Problem of Rule-Following

Frege's rather problematic treatment of fictional thoughts was encapsulated by Geach in the 'Frege Point', which contends that an asserted thought and an unasserted one may share an identical 'kernel'. Such 'kernel' is free of the assertoric force but remains semantically self-contained. I have rejected this idea, since it conflicts with Frege's Original Insight that a content can only be demarcated by the assertoric force; I suggested that we still need to view an unasserted content as 'forced', but with further qualification that distances the force—In short, the assertoric force is *indissociable*, though *distanceable*, from a demarcated content regardless of whether it is asserted or

unasserted.

The indissociability and the distanceability of the assertoric force can be easily confused with each other—just like Frege did: In line with his Original Insight, Frege acknowledges that the assertoric force is indissociable from a truth-conditional content, because otherwise we will have to need another 'force'—such as the 'grasping' of a thought—to demarcate the content. Frege correctly maintains that such a thought-grasping 'force' cannot be really understood as the meaning-demarcating force, since it typically signifies a kind of mental event and does not account for the truth-condition of the content—In the Concept Script, by contrast, what the logicians and philosophers are interested in is exclusively about the content's truth-conditions:

> […] we can probably do without the derivation and explanation of the mental process, if our concern is to decide whether the [mental] process terminates in *justifiably* taking something to be true." (Frege 1918–19, 59)

Thus, although there could well be certain kind of mental process that grasps a thought, it is extrinsic to the demarcation of the meaning of the thought. However, since Frege failed to properly account for the distanceability of the force, he mistook it for the view that the force is nonetheless dissociable from the content—the view that he seems to be also holding—Just as we have seen, he contends that anyone can either assert or unassert, or even just pretend to assert a content. If this should be the case, then it is unclear how an unasserted thought could be demarcated as intelligible at all.

Fregeans might reply that what is indissociable from the truth-conditional content is the assertoric force that demarcates it, while what is dissociable from the content is an agent's additive attitude, which might assent to or dissent from the content. Such a response apparently maintains the essential force–content relation to the extent that the content's truth-condition is still given by the force—Thus it does not espouse a problematic 'force-free kernel' view—Nevertheless, it dissociates the content's truth-condition from the agent's pragmatic condition of using the content. That is, although the assertoric force is retained as essential to the contents, it is made irrelevant to our pragmatic conditions of using them. If so, the force that demarcates a truth-conditional content could be seen as purely formal: It means that it plays no essential role in our practice of using the content in thinking or in speech—while our practice for content-use, in turn, plays no essential role in determining the truth-conditions of the content as well but only indicates a meaning-extrinsic occurrence that expresses certain non-

cognitive dispositions or desires or a mental process that, for instance, 'grasps' the content.

But the pragmatic condition of using a content cannot be reduced to sheer dispositions or mental processes like the grasping of the content; rather, a content's truth-condition and its pragmatic condition of use must be the same.[1] The force that demarcates a truth-conditional content must instantiate the truth-condition as the condition under which the content is used by a token-act of taking it as true—It is true, as Frege states, that our fundamental concern is not about the mental process but how it "terminates in justifiably taking something to be true", but it does not mean that we should regard the force as thereby use-insensitive but only concerns the verbal outcome of our acts (while the outcome is considered as radically semantically autonomous). In other words, it does not mean that we should treat the act of taking *p* to be true and the act of evaluating it as justifiable to take *p* to be true as two different kinds of act:[2] An act that evaluates that it is justifiable to take *p* to be true does nothing other than taking *p* to be true, and vice versa, an act of taking *p* to be true must be inherently an act of evaluating it as justifiable to take *p* to be true.[3] Our act of taking *p* to be true is inherently a knowledgeable rather than sheer dispositional or psychological move.

If this is correct, we may regard Frege's approach as nevertheless leading to a dualism of content and force (insofar that the force must be use-sensitive), each of which is governed by distinct principles of its own kind; namely, it leads to a dualism between the supposedly semantically autonomous content on the one hand and the force answerable to the token-acts of using the content on the other, which is ultimately left inscrutable by the Concept Script.[4]

The dualist problem of dissociating the content's lawfulness from its particular instantiations in use can be compared to the so-called rule-following problem—If we

---

[1] Cf. John McDowell (1976), and (1999).

[2] Cf. Wittgenstein's idea that every possible proposition is legitimately constituted ('Jeder mögliche Satz ist rechtmäßig gebildet.' *TLP* §5.4733). His attempt to closely connect the 'possibility' with the 'legitimacy' of a sentence can be seen as an attempt to incorporate the sentence's use-sensitivity into its evaluability.

[3] The same can also be said of falsehood, i.e. acts of taking *p* to be false. More discussions on disagreements see §5.5.

[4] Cf. McDowell (2005a, 182).

adopt the dualist view, on the one hand it would be hard to find a semantically significant solution to the problem of fiction (since merely observing one's overt verbal outcome can never reach the speaker's real commitment to the content's truth-value); on the other hand it would be also hard to explain how the truth of one's token-act of assertion could be anti-luck (i.e. if the speaker simply 'happened' to have made a correct statement). In a word, it would leave the pragmatic condition of using the content inexplicable, just like—as the rule-following problem shows—that it seems inexplicable whether one's token-expression instantiates or follows a determinate rule or not.[5]

To conclude, the dualist approach needs to be resisted. It means that the pragmatic condition of using a content must be constitutive of the truth-condition of a content, and the agent's token-act of instantiating (be it asserting or distancing) the content's truth is *not* a further act separable from the act of content-demarcation.

2.1.2 Dualism and the Context Principle

It is one thing to claim—reasonably—that there may be no traceable difference in overt expressions between an asserted thought and an unasserted one; it is quite another, however, to claim that an asserted thought and an unasserted one that share the same overt verbal expression are the same thought—just like the Frege Point maintains. I have argued above that adopting the latter claim leads to a force–content dualism.

Rejecting dualism means rejecting that a sentence's truth-condition can be acquired regardless of its context of utterance. A context is the informational background implicit in an utterance, i.e. that which is not in its overt verbal expression. Since according to Frege's Insight there is no fixed expression in our languages that signifies the assertoric force, it is advisable to consider that the force is closely related

[5] Cf. Wittgenstein (1953, §201). Cf. also Saul Kripke (1982), and McDowell (1981), (1984). Note that there is an interesting aspect in Wittgenstein's original example for the rule-following problem, i.e. it exemplifies interpersonal misunderstanding between the performer and the interpreter. This reflects a difficulty in the pragmatic approach to assertion that I shall discuss later in Chapter 6: Although the performer's understanding of the content's truth essentially refers to the conformity to the epistemic standards and norms implicit in the background knowledge of the context of conversation, the epistemic standards and norms related to a conversation are usually *not* symmetrical among its interlocutors.

to the context-sensitivity of meanings.[6]

In fact, Frege was quite ambivalent about context-sensitivity, too: On one hand, he maintains that a proposition as a unity must be closed off by the judgment-stroke;[7] Yet on the other hand, he explicitly put forward a so-called Context Principle in his Introduction to the *Grundlagen*:

> […] never to ask for the meaning of words in isolation, but only in the context of a proposition (*Satzzusammenhang*).[8] (Frege 1884)

Although the expression's overt verbal content such as a word or a phrase has conventional linguistic meaning, its actual truth-conditional meaning depends on the meaning of the proposition expressed as a whole, which varies from context to context. We may call it the *semantic flexibility* of our verbal expressions. Since the consideration of the context typically belongs to the speaker's use of linguistic content in an act of instantiation that decides on whether the content's truth-condition obtains, the Context Principle can be viewed as a claim of the pragmatic significance in determining the meaning of a content. However, we have seen that Frege has de facto committed himself to a force–content dualism, thus it is not surprising that he failed to properly accommodate the Context Principle.

L. J. Cohen has once described the semantic flexibility in terms of the contrast between what he calls 'insulationism' on the one hand and 'interactionism' on the other:

> According to the insulationist account the meaning of any one word that occurs in a particular sentence is insulated against interference from the meaning of any other word in the same sentence. On this view the composition of a sentence resembles the construction of a wall from bricks of different shapes. The result depends on the properties of the parts and the pattern of their combination. But just as each brick has

---

[6] Henceforth in this chapter I take 'context-sensitivity' and 'use-sensitivity' to be synonymous. More discussions cf. Chapter 5.

[7] Cf. Chapter 1 footnote 13.

[8] See Frege (1884). It needs to be pointed out that for Frege the Context Principle is not originally an explicit endorsement of the pragmatics of assertion—It is primarily about names and their quantificational generality (as having variable meanings relative to a parameter) and only then about determining the meaning of a sentence at its context. But as Donald Davidson once remarked: "Frege said that only in the context of a sentence does a word have meaning; in the same vein he might have added that only in the context of the language does a sentence (and therefore a word) have meaning" (Davidson 1966, 22).

> exactly the same shape in every wall or part of a wall to which it is moved, so too each standard sense of a word or phrase is exactly the same in every sentence or part of a sentence in which it occurs. We may sometimes need to look at neighbouring words in order to discover the sense in which a word is functioning in the sentence in question, as we might infer the concavity of one brick from the convexity of its neighbor. But even then the meanings that we discover are not made what they are by one another, any more than the presence of a convex brick actually alters the shape of its neighbour. Rather, the words in the sentence have been given these meanings by diachronic facts of etymology. In the case of an artificial language these meanings have been stipulated or assigned by its inventor […]
>
> Interactionism makes the contradictory assertion: in some sentences in some languages the meaning of a word in a sentence may be determined in part by the word's verbal context in that sentence, though the extent and nature of this determination shows a wide range of variation. On this view the composition of a sentence is more like the construction of a wall from sand-bags of different kinds. Though the size, structure, texture and contents of a sand-bag restrict the range of shapes it can take on, the actual shape it adopts in a particular situation depends to a greater or lesser extent on the shapes adopted by other sand-bags in the wall, and the same sand-bag might take on a somewhat different shape in another wall or in a different position in the same wall. By exploiting local context in this way a language can be much more prolific of semantic variety than insulationism can give it credit for being. Moreover these sense-refinements or sense-modifications are generated by verbal interaction within a particular synchronic state of a language. Many of them, especially metaphors, are not common enough to be recorded in dictionaries or assigned dates and places of origin. Nor are they conceived as independently existing, externally determined constituents of the meanings of the sentences in which they occur, but as arising only within particular sentences through local semantical interactions between the sentence's component words. However, such a sense-modification is sometimes sufficiently useful and occurs sufficiently frequently in actual speech that it comes to be treated as belonging to the relevant word quite independently of context. It is thought of as an alteration in the range of alternative meanings which that word has, and as being available for use in ways that an insulationist account would fit quite adequately. For example, that is how live metaphor passes into dead metaphor. (Cohen 1986, 223f.)

For ‘insulationism’, the context at which a verbal content (a word or a phrase) occurs is irrelevant or only minimally relevant to the meaning of the content, and what is of primary significance is rather the content’s ‘intrinsic’, lexical or linguistic meaning.

Therefore, if there is a pragmatic condition for the language user to instantiate a word or a phrase at its context as meaningful, it can be explained by, and hence reduced to, the semantics of the word or the phrase alone. The difficulty of the insulationist view, however, is not only that it overlooks the semantic flexibility of our language, but also that it remains unclear how those lexical meanings are stipulated in the first place; and consequently, it fails to explain how the conceptual innovation in our discursive practice could be possible—And most gravely, these questions constitute precisely the worry that our acts of using languages fall prey to the problem of rule-following.

The 'interactionism', by contrast, contends that context-sensitivity is constitutive of determining meanings. Although a word has its conventional, linguistic meaning, it does not decide on but only serve as a constraint on the meaning of utterance—In other words, the linguistic convention provides nothing more than contextual, background information that supports the truth-condition of an uttered proposition rather than stipulating its 'literal meaning'.

To sum up, rejecting the dualism means that we need to reject the problematic, 'insulationist' distinction between the speaker's intended meaning and the 'literal meaning' thus spoken.[9] Since our verbal expressions are semantically flexible, the conceptual contents are inherently context-sensitive and inferential instead of being literal. We hence need a rather pragmatist view towards meaning, according to which there is certain irreducibly pragmatic effect that is essential to or constitutive of utterance and hence the truth-condition of the content.

2.2 Semantic Minimalism

The claim that there is an irreducibly pragmatic effect essential to the truth-condition of a content radically differs from the customary idea that the content's truth-condition is fundamentally a semantic issue, in which the pragmatic effect plays, at best, a merely supplementary role—This latter idea is sometimes called *semantic minimalism*. It contends that context, together with other 'non-literal' effects, only has a minimal effect on semantic content (i.e. the truth-conditional content).

Well-known defenses of minimalism include Cappelen and Lepore (2005), King

[9] Cf. Recanati (2004, 4).

and Stanley (2005). Proponents of minimalism do not reject pragmatic effects in language. But they consider them as generally *post-propositional*—namely that the pragmatic effects belong to those that are generally optional and additive, so they do not contribute to the truth-condition of the proposition and its demarcation. For minimalists, the *pre-propositional* pragmatic effects (i.e. pragmatic effects that are constitutive of the proposition's truth-condition) can only be either trivial (i.e. tantamount to semantics) or reducible to a kind of semantics.

There is a more common distinction in studies of pragmatics between the so-called 'far-side pragmatics' on the one hand and the 'near-side pragmatics' on the other. The 'far-side pragmatics' roughly corresponds to the post-propositional pragmatic effects. It generally concerns the speaker's optional enrichment and conversational implicatures, which are carried out by the speaker at will, beyond the truth-conditional constitution of the proposition. The 'near-side pragmatics', in turn, involves filling the 'argument-slots' of the proposition's indexicals, anaphoras, and other substitutive terms—to use Fregean terms, it involves the 'saturation' of the concept with an object—It is thus constitutive of the proposition's truth-condition and is pre-propositional.

According to minimalism, then, although the 'near-side pragmatics' is pre-propositional, it is essentially stipulated by semantics and hence is only 'weakly' pragmatic—It means that the agent has only minimal freedom of having any pragmatic effect on the meaning of the content rather than following the linguistic conventions—By contrast, 'strong' pragmatic effects on the meaning of a proposition may occur only at the 'far-side' of the proposition (hence post-propositionally), which is thus rather derivative from the semantic content and hence is only additive to it.

For example, it can be argued that the Gricean conversational implicature is a post-propositional pragmatic effect—For Grice, indeed, we can say that there are distinctly two steps involved in the interpretation of a proposition: the first step involves the saturation and demarcation of the 'literal meaning' of the proposition (i.e. interpreting what is said), which is essentially a 'sheer' semantic process; while only the second step is the one of 'strong' pragmatic effect, e.g. the conversational implicature (interpreting what is meant by the speaker)—yet it is made *in addition to* the truth-

conditional constitution of the proposition and is always merely optional.[10]

To reject minimalism, we need to see that the speaker's intention is not only constitutive of interpreting what the speaker means by a 'far-side' pragmatic effect but also constitutive of interpreting what is said—There should be no distinct two steps according to which we need to *first* interpret what is literally said through a minimalist semantic process and *then* 'modulate' the 'literal meaning' into an enriched interpretation of what the speaker intended to mean by saying thus and so—In other words, we do not interpret or evaluate an utterance *based on* its 'literal interpretation'. Rather, we need to argue that there is *'strong' and pre-propositional* pragmatic effect constitutive of any truth-conditional content.

How can we interpret a proposition with 'strong' pragmatic effect? A natural answer seems to be that we can carry out such an interpretation regarding the context of the proposition. But it would be problematic to take context as part of the linguistic content. For example, it should be reasonable, following the Context Principle, to adopt the following formulation:

(1) An assertion φ is true iff $\ulcorner \varphi \urcorner$ is true at context c.

And we might regard context c as indication of the pragmatic condition of instantiating the truth-condition of φ. It would be problematic, however, to regard context c and the truth of $\ulcorner \varphi \urcorner$ as standing in an instantiation-relation R, where c is a parameter so that it may or may not instantiate R (e.g. as if treating c as a set of possible worlds). A metalinguistic model for context-sensitivity is of course possible and could be useful, but at the end of the day, such a model does not account for the pragmatic condition of assertion—The pragmatic condition for assertion cannot be reduced to semantics disguised with a pragmatic parameter since the reductive approach tends to be inflationary—This requirement is in line with Frege's Insight: The truth-condition for *S* to assert that snow is white should be the same as the truth-condition of the proposition that snow is white itself.

But what could the pragmatic condition be, if we need to avoid treating the 'strong' pragmatic condition as ultimately expressing a kind of semantic content? It is indeed a challenge to the pre-propositional pragmatics. As for the minimalists, it seems

[10] Cf. Recanati: "according to [the Gricean picture] we process the literal interpretation first, and move on to the derived interpretation only when this is required to make sense of the speaker's utterance." (Recanati 2004, 28).

natural to maintain that the interpretation of 'strong' pragmatic condition hence must be fundamentally different from semantic interpretation, so that the two-step distinction is nevertheless needed: On one hand, the pre-propositional procedure of interpretation is essentially semantic—It is a procedure of 'denotation assignment', in which the interpreter is 'guided' by the context-invariant meaning of the components of a sentence and its syntactic rules[11]—And on the other hand, the essentially pragmatic procedure of interpretation can only be post-propositional and fundamentally irrelevant to the content's truth-condition.[12]

By drawing such a sharp line between semantic and pragmatic interpretation, the minimalists treat the literal linguistic content of a proposition as semantically autonomous in nature. It means that the truth-conditional content, since it is literal, is essentially 'neutral' and 'use-insensitive', while the pragmatics of pre-propositional interpretation, by contrast, is only trivial: It is confined to operations of saturation or instantiation, which is solely governed by the sentence's context-invariant meanings and syntactic rules. However, even if we ignore the 'strong' propositional effects and contexts but only focus on the pre-propositional operation of saturation or instantiation, it remains unclear how the picture of the autonomous, self-interpreting semantic content could explain our acts of language-use—As we have seen in the Fregean dualism, without explanation of this kind we would generally face an enormous challenge to explain how a mental process of language-use that somehow terminates in an utterance coinciding with semantic stipulations and use-insensitive linguistic conventions could justify itself as anti-luck and essentially rule-following.[13]

For minimalists, rules for literal linguistic contents are global and primary in

---

[11] Cf. Jason Stanley (2007, 32).

[12] E.g. cf. Stanley: "[…] semantic interpretation involves the assignment of denotations to elements of a logical form relative to a context, and their combination. Extra-linguistic context enters in only when called upon by a linguistic rule governing an element. The result of semantic interpretation is some kind of non-linguistic entity, such as a proposition or a property, which is then the input to pragmatics." (Stanley 2007, 33)

[13] Emma Borg (2004) suggests that we should allow pragmatics to play a non-trivial but integral role in the pre-propositional determination of the truth-condition of a content; Yet she takes the pragmatics to be nonetheless belonging to a "modular theory of mind" that does not conflict with semantic minimalism but can cooperate with a formal semantic theory of meaning. Cf. Borg (2004, 60). But it is unclear how the cooperation could be possible under such a semantic/pragmatic distinction. Cf. also Recanati (2010, 2).

terms of explaining any content's truth-condition—they are just out there at our disposal—while the pragmatics of language-use is only local and secondary to the content's truth-condition. But I think it should be the other way around: the pragmatics of language-use should be global and primary in terms of explaining an utterance's truth-condition, while rules for literal linguistic contents are rather local and dependent on the truth-condition specified by the context that brings relevant linguistic rules into being authoritative. There is no status for a content to be absolutely literal—even a formula like 1+1=2 does not enjoy absolute context-insensitivity but asks for relevant knowledge about natural numbers, numeral systems and notations etc. that is not explained by the literal content of the formula itself. In short, since there is no one-off translation that can convert any context-sensitive sentence into a context-free one, a global account of truth-conditional content cannot do away with context-sensitivity.

2.3 Essential Context-Sensitivity

The minimalists contend that pre-propositional pragmatic effects are restricted to operations of saturation or instantiation, which are governed by, and hence ultimately explained by, purely semantic interpretations. But if we insist on the context sensitivity of contents, it seems that there is no literal, context-invariant meaning available for us to easily establish the truth-condition of a content in terms of semantic interpretation. Indeed, as I will show in §2.3.1, our linguistic expressions are generally equivocal and essentially context-sensitive. In §2.3.2 I will discuss Quine's treatment of singular terms and his holistic approach to meaning, which correctly addressed the overall context-sensitivity of truth-conditional contents. Yet the Quinean holism, as insightful as it is, overlooks the 'intralinguistic' significance of the pragmatic condition of the use of a linguistic content at its context, which, as I will suggest in §2.3.3, must be constitutive of the content's truth-condition.

2.3.1 Context-Sensitive Expressions

The most straightforward example of context-sensitive expressions are indexicals. Although every competent English speaker understands what the word 'this' lexically means, we always need relevant information in corresponding context to determine what the speaker is actually referring to by using that word. The indexicals as such,

therefore, are equivocal—Used in a context, the word 'this' may refer to a book, while in another context it might refer to a house instead—Since the meaning of indexicals always depends on their token-expressions in different contexts, they are also called *token-reflexive expressions*.

The thorny problem of token-reflexive expressions is that we cannot interpret such an expression based on purely linguistic explanation alone—As Aristotle once described the problem:

> "How will you prove what a thing is? For it is necessary for anyone who knows what a man or anything else is to know too that it is. […] But if you are to prove what it is and that it is, how will you prove them by the same argument? For both the definition and the demonstration make one thing clear; but what a man is and that a man is are different. […] So when you define what it is, what will you prove? Triangle? Then you will know by definition what it is, but you will not know if it is. But that is impossible." (92b4-18)[14]

This passage raises the question in a similar fashion like the rule-following problem: Namely, it is unclear how we can account for the relation between a term's linguistic definition and its use in a token-act of demonstration (i.e. in a token-reflexive expression) without falling into a circular explanation.

The word 'this', more accurately speaking, is a demonstrative. The category of token-reflexive expressions (i.e. indexicals) is allegedly wider than that of demonstratives, due to Kaplan's distinction between 'pure indexicals' (such as 'I', or 'today') and 'true demonstratives' (such as 'this', 'that'). I will not debate on this distinction here but only speak generally of token-reflective expressions—For Kaplan, the equivocality of 'pure indexicals' can be resolved by introducing further, more accurate, linguistic explanations: e.g., it is believed that the equivocality of 'I' or 'today' occurring in an expression can be eliminated by substituting the indexical with a proper name (e.g. 'YZ') or a definite time (e.g. '1.1.2015')—In other words, the solution to its equivocality is believed to be able to take place at purely linguistic level, while the solution for 'true demonstratives', by contrast, is argued to be requiring an extra bit of intention from the speaker.

However, I do not think there can be a purely linguistic solution even to the determination of the meaning of 'pure indexicals'. For example, it might be argued that

[14] Aristotle. *Analytica Posteriora.* II 7, English translation in Barnes (1984).

sentences like "The current French king is worried" may have its indexical 'current' determined by a date, say, when the sentence is written in a diary dated 1792, which makes the sentence sensical. Yet imagine that some archeologists found out later that that diary, though dated 1792, was actually written during the time of Caesar, then we can by no means say that the sentence about the French king, even taken the whole literal content from that diary into consideration, is sensical.[15]

In fact, context-sensitivity is ubiquitous, and there is no simple linguistic solution available to the determination of equivocal, context-sensitive terms. Suppose John says: "Sophie is a cyclist." And I ask him: "What do you mean she is a cyclist?" In a sense, my question can be regarded as a request for purely linguistic explanation—I expect further information to explain the equivocal 'cyclist'. (Does my interlocutor mean that Sophie rides a bicycle to work? Or does he mean that she cycles for exercise or even as a profession?)—Since this example does not concern any explicitly token-reflexive expression, it might seem that we can inquire into the lexical meaning of an expression by simply substituting it with a more accurate and rather univocal one, e.g. that Sophie is a cyclist in the sense that she commutes by bike—Wouldn't that be a simple linguistic solution?

We need to be more careful here. To begin with, the sentence '*x* is a cyclist' might be generally analyzed as follows (given that the word 'cyclist' is equivocal):

(1) $\lambda x\, \lambda \boldsymbol{X}$ [(CYCLIST IN THE SENSE OF $\boldsymbol{X}$) (x)]

In the conversation where John says "Sophie is a cyclist", suppose the referent of the name 'Sophie' is univocal among the speaker and the hearer so that we can substitute *x* with 'Sophie', then we have, apparently, a simple equivocal expression:

(2) $\lambda \boldsymbol{X}$ [(CYCLIST IN THE SENSE OF $\boldsymbol{X}$) (Sophie)]

However, if Sophie commutes by bike, we will actually have:

(3) $\lambda \boldsymbol{X}$ [(CYCLIST IN THE SENSE OF $\boldsymbol{X}$) (Sophie) & (COMMUTER) ($\boldsymbol{X}$)]

We cannot equate 'commuter' with 'cyclist' since they are not overlapping but only

[15] I will have more discussion of the relation between context and token-reflexive expressions in §2.4. For more discussions on the irreducibility of the first-person cf. §4.3, further discussions on its relation to proper names cf. §5.3.

occasionally intersecting.[16] Thus the only hope, it seems, to eliminate the variable ***x*** is not substituting any further linguistic explanation by the term 'cyclist' but introducing *another assertion* "Sophie is a commuter", so that the intersection of the two assertions is held by the name 'Sophie':

(4) (CYCLIST) (Sophie) & (COMMUTER) (Sophie)

The analysis shows that even for sentences that do not contain token-reflexive expressions, the elimination of equivocals cannot be done by substituting parts of a sentence at subsentential level. Rather, it can only be carried out at the level of propositions.

But how can we explain the semantic stability of a proper name among its occurrences in multiple assertions? It might be suggested that proper names (such as 'Santa Claus') and definite descriptions (such as 'the first atom in the universe') can secure their referents to determinate individuals regardless of their contexts. But there is no reason to view any of such contents as having absolute, context-invariant meaning.[17] After all, even a proper name is fundamentally equivocal in our conversations: 'Socrates' may refer either to a historical figure, or his corpse, or his thought, or a sculpture of him, or simply another individual who happens to bear this name, etc. Therefore, we need to explain how one can know the two tokens of the name 'Sophie' occurring in the two assertions as having the same referent.

The meanings of proper names or definite descriptions are nevertheless context-sensitive. And just like token-reflexive expressions, it seems that there is no purely linguistic explanation available to settle their meanings. In sum, proper names, definite descriptions, and token-reflexive expressions are called singular terms, and it raises a big question how we can account for the context-sensitivity of our use of all the singular terms. In the next subsection, I will briefly discuss Quine's treatment of singular terms, his recognition of the ubiquity of context-sensitivity, and his commitment to a holistic view towards meaning.

---

[16] More discussions cf. Quine (1960, §41).

[17] It is another issue that a proper name, as Kripke maintains, is a rigid designator. I cannot discuss the problem of rigidity here, but as I will at least indicate in §2.4.1, the idea of rigid designator in general does not square with context-sensitivity very well.

2.3.2 Quinean Holism

With his famous slogan "to be is to be the value of a variable", Quine tries to explain reference and hence the use of singular terms away into a web of beliefs. A crucial idea for Quine is that variable must be bound by a relative clause which can be formulated as '*x* such that *F*(*x*)'[18]—For him, the variable is the 'figure in the programming', 'the focus of indecision'; and to identify an object in reality, we need to make 'cross-reference' to indefinitely many previous predications—It does not matter, therefore, whether there is a literal meaning of 'Phosphorus' that determines our use of the term, or whether it may or may not be different from the literal meaning of 'Hesperus'. All that matters, instead, is whether the relative clauses in which the terms occur in our use intersect each other: "we shall want usually to interpret '*F*', '*G*', etc. not just by substitution of particular words or phrases such as 'man' or 'Greek' or 'white whale', but by substitution of […] relative clause".[19]

Such a treatment of singular terms through variables is a plea for thoroughgoing context-sensitivity. For Quine, the semantic value of a singular term can only be determined according to the relevant clauses in which it occurs, while the 'radical translations' and thus the interpretations through literal linguistic meaning are only possible "within the terms of some particular system of analytical hypotheses."[20] As a result of the recognition of overall context-sensitivity, Quine concludes that the unity of thought can only be acquired as a coherent and holistic network of our best scientific theories and beliefs at the time.[21]

Nevertheless, for Quine the 'web of belief' or 'web of theories' relies on a

---

[18] Quine begins a lecture on the variables as follows: "The variable qua variable, the variable an und für sich and par excellence, is the bindable objectual variable. It is the essence of ontological idiom, the essence of the referential idiom." (Quine 1972).

[19] Quine (1972). Quine admits that the bound variable cannot be confused with freely substitutional variable. The bound variable is objectual, while object means that which can be referred to as values. But a variable is void of value in itself. It is hence actually confusing how we, following Quine's advice, could avoid confusing it with the substitutional variable.

[20] See Quine (1960, §16).

[21] "The idea of defining a symbol in use was, as remarked, an advance over the impossible term-by-term empiricism of Locke and Hume. The statement, rather than the term, came with Frege to be recognized as the unit accountable to an empiricist critique. But what I am now urging is that even in taking the statement as unit we have drawn our grid too finely. The unit of empirical significance is the whole of science." (Quine 1951).

bottom-up fashion of saturation through observation: For him, theory is classification or differentiation at higher orders. The higher it is, the more general and stable it becomes. But the ultimate source of verification of our beliefs comes not from generalized theories but from the periphery observation of the ever-changing empirical intake, which we express in occasional observation sentences. The observation sentences, therefore, are true on some occasions but false on others. But if the truth-conditions of such observation sentences do not come from theories, how can we understand its being constrained by the relative contexts? It is not clear how Quine's idea of the semantically autonomous holistic web of theories could be compatible with his idea of the bottom-up fashion of saturation through observation.

If we adopt the reduction of singular terms to bound variables and Quine's treatment of occasional observation sentences, we would be only capable of referring to *hyper-finely individuated object-slices*: If Sophie was seen seated but now standing, the expressions "Sophie was seated" and "Sophie is standing" do not co-refer to a single entity whose property has undergone a change; rather, they refer to two different entities: The seated Sophie_1 ceased to exist and a new Sophie_2 comes to existence.[22] Sure, both Sophie_1 and Sophie_2 are bound by their relative clauses, but it does not explain how they can be grasped as anything other than discrete slices. Since the occasions for observation are constantly changing, we can never really refer to a cat but always a cat-slice at t_1 and another cat-slice at t_2 … and so on. Also, an agglomeration of such cat-slices does not add up to a cat—A cat, as an object, differs from a cat-slice in that it must be able to be reified over time and in changing circumstances.[23]

In his "Identity and Predication", Evans explicitly points out this problem: Namely, it is unhelpful to believe that when you point to a rabbit—

[22] I borrowed this example from Shields (1999,159). Cf. also Quine (1960, §41).

[23] The discussion at this point could become rather complicated if we concern the relation between time and proposition, e.g. concern the debate over eternalism and temporalism. I cannot take on the debate here but only provide a very primitive observation about the infeasibility of hyper-finely individuated object-slices. Cf. Rödl's discussion on endurance and perdurance in idem (2012, 97f.). I will also touch upon some issues concerning temporality in discussing the dynamics of belief-change in §6.2—We will be able to see that asserting (an observation sentence) is not a lifeless repetitive instantiation of literal contents but expression of knowledge, which implies consciousness of change from not-knowing to knowing—as an inferential movement of integrating reasons.

> […] you have thereby pointed to many (on some reckonings, infinitely many, but anyway, many) parts and stages. If 'this' is to be treated as a demonstrative singular term, and 'rabbit' is to divide its reference over rabbit parts, we must suppose some fixed principles for dividing everything into parts—as upon a butcher's wall chart. For dividing things into stages we need something more complicated—a fixed point of origin, say the birth of Christ, and a determinative length of time, say a day. But these desperate manoeuvrings serve only to re-emphasize how far we are being pushed into assignments of utterly unempirical dispositions. Midnight turns out to be a magic moment, semantically speaking; how odd that the language does not have the resource to voice the significance of its passing. (Evans 1985a, 46)

If our use of demonstratives is ultimately explained or saturated by sheer occasional observations, it would be unclear how they can be semantically stabilized and participate in the web of occasion-neutral theories through cross-reference. In his later years, Quine tried to explain the stability of observation sentences in behavioristic terms such as the stimulus–response model, according to which the repetitious stimuli of cat-slices exemplify an inertia-like resemblance in perception—like the persistence of vision. He has expressed such an idea on many occasions in his late writings. The following passage is quite representative, where he speaks of a so-called 'preestablished harmony':

> The preestablished harmony is needed to account for our meeting of minds not only on aardvarks, but also on what to call them: on the mellifluous Dutch disyllable 'aardvark' itself. The phonetic constancy of a word, from one utterance of it to another, is itself a product of the speakers' subjective standards of perceptual similarity. Thanks to the harmony, communication proceeds apace. Oh, we sound alike. Oh, who says so? Each of us, by his own standards of perceptual similarity, all of which are in harmony. (Quine 2000)

Through the presumption of harmonic resemblances among unrepeatable qualia and stimuli, Quine concludes that the very basic way for us to stabilize a semantic content is through our natural instinct or certain innate capacity. But it makes the content's truth-condition inexplicable, because it leaves no room for the possibility of specifying the occasion of saturation in an observation sentence other than in behavioristic terms.

Although Quine has pointed out, quite accurately, that the context-sensitivity of our beliefs and propositions is rather ubiquitous, he mistakenly based his holism on the idea that we can have a web of semantically autonomous contents—To clarify this

point, we can interpret Quine's idea in this way: For him, though a singular content does not have any 'absolute' literal meaning, it may well have an infinite set of possible literal meanings, and the context is just an indication of which of the possible meanings should be picked out at the relevant occasion. But according to our discussion above, Quine's treatment of the saturation of observation sentences leads to a dualism similar to Frege's with its consequent rule-following problem: On one hand, although we can give a literal, semantic interpretation of a demonstrative by substituting it with a binding clause in an observation sentence, it seems never able to really pick out an occasion-neutral object rather than introducing further occasion-dependent object-slices; On the other hand, the stability of a cross-referred content is explained by pragmatic interpretations in behavioristic and hence non-truth-conditional terms—The token-act of uttering an observation sentence is not capable of understanding itself as rendering any privileged meaningful use of linguistic content that exemplifies certain rules—In other words, according to such a dualism we would find ourselves desperately oscillating between an autonomously self-interpreting world of meanings as a whole thinking in place of us on the one hand and a world of mere sentient behaviors on the other.[24]

2.3.3 Essential Pragmatic Condition

Although the Quinean holism correctly acknowledges the overall context-sensitivity of truth-conditional content, the pragmatic condition of interpreting a content under its relevant occasion is given by a behavioristic approach. We have seen that it cannot be satisfactory to assume the continuous existence of an object, say, cat, and the semantic stability of the use of the term 'cat' to be nothing but a visual perception of what appears to be a cat with co-occurring phonetic perception of what sounds like the word 'cat'—Such a behavioristic approach would dispense with the comprehensibility of meaning in our speech acts all together—In fact, the context-sensitivity cannot be explained away, or so I will argue, using any 'external' approach to the pragmatics of interpretation (i.e. treating the pragmatic condition as not constitutive of the truth-condition of the content).

According to Evans, the fundamental problem of the apparent indeterminacy of token-reflexive expressions consists in the presumption that there is a purely linguistic

[24] Cf. McDowell (1987, 97).

approach to explain away equivocal expressions on the one hand, and a purely extralinguistic and non-conceptual approach of perceiving objects on the other. Although with linguistic tools we can adjust the verbal accuracy of an expression, say, changing from saying that what is pointed to is a rabbit to saying that what is pointed to is a bunch of rabbit-parts, such a purely linguistic approach alone cannot really render the meaning of a token-reflexive expression as determinate. To be able to determinately interpret and evaluate a token-reflexive expression, we need to regard the pragmatic condition of using the expression as constitutive of the condition for determining its meaning—Evans calls this requirement Russell's Principle, which he formulates as follows: "a subject cannot make a judgment about something unless he knows which object his judgment is about".[25]

In a sense, this may seem like a principle about one's extralinguistic mental property. For example, Evans speaks more specifically of a knowledgeable condition—which he calls *discriminating knowledge*—constitutive of our perceptual judgment to distinguish the object from all other things, i.e. to single out the boundaries of our demonstrative identification. However, upon careful reading we should be able to see that this is not a condition for visual discrimination but for knowledge; and the Russell's Principle is rather a principle about the nature of judgment and the truth-condition of contents: First, Russell's Principle maintains that we cannot make a judgment about something if that thing is not determinately referred to as an object in that very judgment. Since, as discussed, determinately speaking of an object in an empirical sentence generally requires the use of singular terms, it means that the use of singular terms cannot be explained away but constitutive of the truth-condition of the expression. Second, the constitutive condition specified by Russell's Principle is a pragmatic one, because it asks for the subject's intention and knowledge implicit in the judgment.[26] From Russell's Principle, therefore, we can conclude that the pragmatic condition of using or interpreting a content cannot be explained from outside the content's truth-condition.

---

[25] Evans (1982, 89).

[26] Admittedly, we can formulate other conditions for making judgments beside the condition formulated by Russell's Principle: e.g. a subject cannot make a judgment unless he is alive, or awake, etc. But these conditions are not sufficiently specific about what is pragmatically constitutive of a content, namely the subject's knowledge that its truth-condition obtains. In fact, we can see that Quine's behavioristic formulation has exactly the problem.

It might seem that Russell's Principle does not apply to generic statements such as "Snow is white", since such statements do not contain any singular terms. But they certainly have the similar context-sensitivity as those with explicit singular terms have—For instance, the term 'snow' could be seen as having an implicit classifier, so that when we talk of snow we actually talk of, say, heaps of snow.[27] But sometimes we might also speak of snow under some other sortal concepts, e.g. a thin layer of snow, or particles of snow crystals smaller than the wavelength of visible light (so that it would not be white). Or, maybe we are speaking of a symbol of snowflake (which could be drawn in any color), or a heap of snow under a different lightening condition, etc.—These examples indicate that purely linguistic explanation also cannot account for the truth-condition of generic statements, which always depends on specific discursive contexts. As a generic statement, the meaning of the sentence "Snow is white" actually requires more complex 'discriminating knowledge' than a sentence of direct demonstration such as "This is white" does.

Our discussion of Russell's Principle shows that the problem of context-sensitive meanings, at truth-conditional level, is more a problem about the speaker's intention and background knowledge than a problem about indexicality and saturation. Nonetheless, the relationship between context and indexical expressions needs to be sorted out more carefully: E.g. according to Kaplan's influential theory of indexicals, it is one thing that a speaker has knowledge implicit in the pragmatic condition for expressing a context-sensitive, truth-conditional content, and it is quite another that the truth-conditional content's indexical terms are saturated in assertions. Adopting such an approach, however, would be mistaken—as I shall discuss in the following section.

2.4 Pragmatics of Assertion: Three Models

I have shown in the previous section that our linguistic expressions are essentially context-sensitive. This idea opposes the minimalists approach to pragmatics: As mentioned, the minimalists consider the near-side pragmatic condition as only weakly pragmatic. It involves, for instance, saturations and interpretations of indexicals in a content, which is ultimately governed by a process of semantic interpretation. And once

[27] Cf. Strawson (1959, 207).

the content is saturated, it will be out there at our disposal, with its truth-condition intact. In other words, the content will be regarded by minimalists as semantically 'neutral' and self-contained; It might be put in different contexts of use and have different interpretations, but such context-sensitive interpretation is only a post-propositional 'bonus effect' of far-side pragmatics, which is *not* constitutive of the content's truth-condition—Opposing the minimalist approach, I have suggested that the 'strong' pragmatic condition of context-sensitive meanings cannot be explained from outside the content's truth-condition; rather, we need to argue that the pragmatic condition is pre-propositional, i.e., constitutive of the content's truth-condition.

There are multiple possible approaches to account for the essential pragmatic condition of context-sensitivity at this point. In this section, I will examine three influential approaches: the Kaplanian, the Lewisian, and the Austinian ones, respectively. I will argue that only the last, i.e., Austinian approach is feasible.

2.4.1 The Kaplanian View

Based on his theory of indexicals, Kaplan's approach to the pragmatic condition of context-sensitivity is perhaps the most influential. In arguing for the essential pragmatic condition of context-sensitivity, he nevertheless maintains that a content's 'literal' (i.e. use-insensitive) semantic interpretation of saturation and its use-sensitive, pragmatic interpretation are different in kind. The difference of Kaplan's from the minimalists' approach consists only in the sense that unlike the minimalists, who put the use-sensitive pragmatic interpretation after the saturation of a truth-conditional content, Kaplan put the use-sensitive pragmatic interpretation before the saturation of a truth-conditional content. In this subsection, I argue that such an approach is not really different from semantic minimalism.

For Kaplan, therefore, it involves two stages to assert a content: the first is a pragmatic interpretation according to the content's context, and the second is a semantic interpretation according to the circumstance at which the content is saturated. These two stages of interpretation are depicted by Kaplan through two functions: The first is a function from contexts to contents, which Kaplan calls *character*, which roughly corresponds to the linguistic expressions of a sentence; an interpretation of the character of a sentence at its context yields a truth-conditional content, which, in turn, serves as a second function from different *circumstances* of evaluation to assertions. (If we regard a

content's circumstance as indicating a possible world, we can see the first-stage interpretation as determining the content's intension, while the second-stage interpretation as determining its extension.)[28] We can illustrate the two stages roughly as follows:

For Kaplan, a context is an abstract object, which is a quadruple of an agent, location, time and world: $\langle c_A, c_P, c_T, c_W \rangle$. For example, the character of 'I' at the context where Sophie speaks of herself has the name Sophie as the value of c_A. Through the interpretation of the character at its context of utterance, we can hence determine what is said by the speaker. However, knowing what is said by the speaker is not yet determining whether the speaker says something true: It is one thing to determine the c_W at context c in which the speaker says "It is snowing", so that we may understand that when the speaker utters the character 'it' she is speaking of, say, Berlin. But it is another to determine the world w at which we may evaluate whether the content put forward by the speaker is true, i.e., whether Berlin is really snowing—This hence requires a 'double indexing': the truth-value of a sentence is relativized at two stages: the context of utterance on the one hand, and the world or the circumstance of evaluation on the other—While the first stage eliminates misidentification, the second stage examines whether the predication applies.[29]

We evaluate a speaker's utterance 'It is snowing' based on the object or the circumstance that the speaker is demonstrating in saying so. However, according to Kaplan's two-stage division the interpretation of the indexical in evaluation is independent of the context of utterance: It behaves rather like a rigid designator.[30] But if

[28] The basic idea of Kaplan's theory of indexical can be found in idem (1989a) and (1989b).

[29] I will have more discussion on this in §5.4.3, where I discuss the so-called 'two-dimensional semantics' initiated by Stalnaker.

[30] The Kaplanian semantics can speak of a rigid designator relative to a context: For example, 'I' rigidly designate a context-relative agent c_A, i.e. it rigidly designates Sophie if Sophie is the speaker at given context. But such kind of context-relativity does not really yield any context-sensitive interpretations. A famous counterexample is that supposing Kaplan was standing in front of a mirror and said to himself: "I am David Kaplan, you are not David Kaplan", in context-sensitive interpretations this can be explained by the equivocality of the name 'David Kaplan' in that sentence, but according to the two-stage picture, the sentence must be evaluated

so, how does it have anything to do with the speaker's intention in demonstration? In fact, just like the minimalism, Kaplan's approach maintains that the saturation of a content is essentially governed by semantic interpretations of the context-invariant, 'literal' meanings of the content. The essential pragmatic process of interpretation, by contrast, only fixes the values of characters at subsentential, 'preparatory' level: E.g. it substitutes 'it' with 'Berlin' and then its job is done. As to whether the sentence 'it is snowing' as a whole may or may not express a truth, the pragmatic condition simply has nothing to say. Therefore, even though Kaplan's treatment of the pragmatic condition of context-sensitive meanings is apparently pre-propositional, it is nonetheless not constitutive of the content's truth-condition—just like the minimalism.

Kaplan is correct in maintaining that in making an assertion or saturating a content, the indexical should be considered as referring to an object but not a context. To evaluate the sentence 'It is snowing' we only need to see whether it expresses the truth. The truth could be about a specific time and place, i.e., have them as part of its truth-condition, so that we can say that a content is true at a specific time and place; but it does not makes sense to say that an assertion is true only at a specific time and place, e.g. at the time and place of utterance—The time and place at which we evaluate 'It is snowing' need not be the time and place of the context at which the speaker uttered the sentence—This is the central characteristic of the Fregean thought: namely, I may say 'it was snowing yesterday' to express the same thought as that which I expressed yesterday by saying 'it is snowing today'. For Frege, the difference between the two sentences consists only in the wording, and we thus need a semantic theory to explain how the two expressions with different linguistic contents have the same truth-value. For this reason, Kaplan sees it necessary to make the circumstance of evaluation independent from the context of utterance. This makes easier to explain compositional sentences like the following:

(5) I met John yesterday, he wouldn't believe that it is snowing now.

But I think it is misleading to claim that the semantic meaning of the indexical 'now' in

---

at the second, context-insensitive stage, so that it can only receive the rigid designator interpretation, according to which the sentence is rendered contradictory. In general, whether a term is a rigid designator should not concern our treatment of essential context-sensitive expressions. (E.g. consider a similar example with proper names: 'Marx is dead, long live Marx!'). Cf. also Evans (1982, 61ff.).

the indirect speech is independent from the context of original utterance. Sure, in reporting John's speech I changed his word from 'tomorrow' to 'now'—The word 'now' is not only an indication of the circumstance of my own assertion but also indication of me keeping track of the truth-condition of John's original statement, so that we have to find a way to fix the indexical to make the indirect speech—However, as I will argue in more detail in §§5.4–5.5, what we need to do is rather to fix (to preserve) the context of utterance in evaluation—i.e. denying that context can be freely shiftable in evaluation—rather than fixing the indexical and making it context-independent. The intelligibility of the circumstances' change from 'tomorrow' to 'now' is derivative of, rather than independent from, the acknowledgement of the utterer's pragmatic condition of asserting that a content's truth-condition obtains at its context.

2.4.2 The Lewisian View

Contrary to Kaplan, Lewis correctly observed that a sentence's circumstance or 'index' must be context-dependent. We can illustrate the way the index depends on the content's context roughly as follows:[31]

(*L-model*) SENTENCE —*index*→ CONTENT —*context*→ ASSERTION

According to Lewis, a sentence is an n-tuple such as ⟨'it', 'is', 'snowing'⟩, and it may be used at different circumstances indicated by, e.g., world and time. From the sentence together with its indices we can have a circumstance-sensitive content, e.g.: ⌜ it is snowing. ⌝ at ⟨w, t⟩. The content, in turn, is essentially relativized by context—It is a function from context to truth-values—Therefore, the context is constitutive of the content's truth-condition.

Since a sentence's index or circumstance must be context-dependent, the world and time must be fixed by the context of uttering or asserting the content. Therefore, for Lewis, indices are features of the context—We can thus write them instead as w_c and t_c, and the truth-condition of a sentence can thus be simply put as:

(*L*) A sentence φ is true iff ⌜ φ ⌝ is true at context c, ⟨w_c, t_c⟩.

[31] Cf. Lewis (1980). Though Lewis himself did not use the term 'assertion' but only speaks of 'semantic values' instead.

Since indices are just features of the context, there are not really two stages in the interpretation of a sentence—The semantic interpretation and the pragmatic interpretation are not really different in kind. The indexicals that are conceived by Kaplan as fixed and rigidly designating are now explained by the context of the speaker's utterance.

In doing so, the Lewisian content is radically contextualized. It is the speaker's context alone that directs the truth-condition of contents. Although Lewis also implemented the 'double indexing' in his treatment distinguishing world-dependence from context-dependence, the set of worlds at which a sentence is evaluated must be essentially the set of worlds at which the speaker locates herself in, which Lewis calls 'centered worlds'. To interpret a content uttered by a speaker is to view it as essentially the speaker's self-locating content, i.e. the speaker's doxastic content; and vice versa, for someone to believe a content is to self-ascribe the content as a property of hers.

The idea of self-locating content can be traced back to Hector-Neri Casteñeda's discovery of what he called 'quasi-indexicals' (or 'quasi-indicators')—Namely, there seems to be an ineliminable expression of the first-person in a content of belief.[32] John Perry developed this idea into a theory of indexicals in the late 1970s, which has found much resonance in Lewis's work.[33] Since according to Lewis, as we have seen, all beliefs are self-locating contents and hence are properties of the speakers, they radically differ from the orthodox Fregean thoughts: For Frege, the central characteristic of the content in the Concept Script is that it must be intellectually sharable. But for Lewis, the content is reduced to the context as a property 'privatized' by the speaker.[34] The problem of such Lewisian approach is that it violates Frege's Insight that the assertoric force does not alter or make extra contribution to the content—Hence it seems impossible for such a content to receive interpretation as a Fregean thought does—By insisting the ineliminable indexical, the speaker's self-locating content is made inscrutable.[35]

---

[32] I will discuss Casteñeda's 'quasi-indexical' in §4.3.2.

[33] Cf. Perry (1977), (1979), and Lewis (1979).

[34] Another extreme at this point is perhaps Stalnaker, who reduces context to a set of contents, which he calls the 'common grounds', cf. my discussion in §5.4.3.

[35] Frege himself also apparently struggled to account for the first-person and demonstrative thought. E.g.: "Now everyone is presented to himself in a particular and original way, in which he is presented to no one else. When Dr. Lauben thinks that he has been wounded, he will

I agree with Lewis that indices are features of context-dependency; namely, the use of indices involves complex contextual information on which the speaker is authoritative. But we do not need to defend the context-dependency in the way Lewis (and Perry) did. He failed to realize that the ineliminable expression of the first-person and of the speaker's background knowledge is not a kind of semantic property, but an essentially pragmatic one.[36] Lewis is correct in maintaining that the context of utterance and the context of evaluation are essentially the same—It is not sensical to evaluate a content from a shifted context—But he failed to explain how we can thereby avoid solipsism. (More discussion on Lewisian contextualism in §5.4.1.) There must be a way to compare and evaluate contents across different contexts other than context-shift (which, as I will suggest in §5.5, is a procedure of context-revision), otherwise evaluating content across contexts would be impossible.

2.4.3 The Austinian View

The consequence of the Lewisian context-prioritizing proposal is that while facing a sentence we must always go *top-down* rather than *bottom-up* to catch its meaning. I.e., we need to start from the whole sentence as an assertion and then analyze the content with respect to its context, rather than the other way around; and the speaker's pragmatic condition of making the assertion is thus always constitutive of the content's truth-condition. This top-down pragmatic view actually coincides with Frege's idea of the 'assertoric sentence-form' (*die Form des Behauptungssatzes*):

> We express acknowledgement of truth in the form of assertoric sentence. We do not need the word 'true' for this. And even when we do use it the properly assertoric force does not lie in it, but in the assertoric sentence-form; and where this form loses its assertoric force the word 'true' cannot put it back again. (Frege 1918–19, 63)

The Kaplanian view hold a bottom-up view towards pre-propositional pragmatics:[37] As indicated, for Kaplan, the pragmatic interpretation handles the context-sensitive

---

probably take as a basis this original way in which he is presented to himself. And only Dr. Lauben himself can grasp thoughts determined in this way." (Frege 1918–19, 66).

[36] Robert Stalnaker seems to be pointing towards this direction in his late writings cf. idem (2014).

[37] On the distinction between bottom-up and top-down pragmatics, cf. Recanati (2010) 'Introduction' and Chap. 1. Cf. also idem (2007) Part 1 & 4.

variables and equivocals of a sentence at the first, lower stage; but it does not affect the truth-condition of the sentence as a whole, which is only subject to use-insensitive semantic interpretation according to circumstances at the second, higher stage. A top-down pragmatic view, quite the contrary, needs to reject the distinction between semantic and pragmatic interpretation of a content. The pragmatic condition does not operate merely at the component-level of a sentence; rather, it inherently operates at the assertion-level. Since to interpret a content is to make an assertion on its truth, rejecting the distinction means that the assertion must put its content's use-sensitivity in order. Lewis partially adopted this view, but his treatment of context failed to properly address the significance of the speaker's pragmatic condition. By contrast, what is crucial here in emphasizing the context-sensitivity of contents is not the relativization of token-reflexive expressions by contexts as such, but an indication of the dynamics of contents used in our acts of speech[38]—We need a full-blown, top-down pragmatic treatment of assertion capable of taking the speaker's intention and background knowledge into account. I would like to suggest an Austinian model as a suitable candidate for this.

The model is based on J. L. Austin's speech act theory. According to Austin, every use and interpretation of linguistic expressions is a speech act. A speech act is not only the act of expressing a linguistic content—which Austin called 'locutionary act', e.g. the phonetic act of saying thus and so or the act of repeating a sentence as such, etc.—but more importantly, it is the act of what one does *in* saying thus and so, which Austin called 'illocutionary act' (usually understood as the speaker's intention). For example, if my boss says to me: "you are fired", he is not just communicating the locutionary content that can be used to describe *a* situation of firing; by uttering the locutionary content he is actually performing *the* act of firing. A more precise discussion of the nature of Austinian illocutionary act must be postponed to Chapter 4. Here I would like to simply lay my cards on the table: I suggest that we can treat assertion in the Austinian way, i.e. as an illocutionary act—Because Austin argues that the illocutionary act is explanatorily primary, so that to understand a speech act is inherently to understand it as an illocutionary act, we can thus finally depict the relation between linguistic sentence, truth-conditional content, and assertion in a top-down fashion, roughly as follows:

[38] Cf. Dummett (1981, 382ff.).

(*A-model*) SENTENCE ⟨*distancing*$_2$⟩ CONTENT ⟨*distancing*$_1$⟩ ASSERTION

The 'content' roughly corresponds to the reported content in an indirect speech, which according to Austin is presented by a locutionary act. (More accurately speaking, it indicates a 'rhetic act'. See my discussion in Chapter 4.) To utter an indirect speech, one needs to create a distance in the context at which she asserts (marked above as 'distancing$_1$'). The 'sentence', in turn, roughly corresponds to the locutionary expression which Austin called 'phatic act'. Examples of phatic acts include 'direct speech report', fictional display of speeches, etc. (See §5.5.1.) In those cases, the context becomes further distanced so that there is a shift in context at local level (marked above as 'distancing$_2$'). Since according to the Austinian model, the locutionary contents cannot stand alone but are only derivative of the utterer's illocution, we have a picture that is pragmatic all the way down.

2.5 Assertion as Performative

In this chapter, I discussed Frege's dualism and its difficulty in accounting for the problem of rule-following. If we take the rule-following problem serious enough, we have to see that an agent's pragmatic force must be constitutive of the truth-condition of the content uttered by the agent. In this regard, I rejected semantic minimalism, Quinean behaviorism, and the Kaplanian two-stage picture. These pictures generally dissociate the agent's pragmatic condition from the assertion's semantic truth and thus fail to properly account for the essential pragmatic condition, which must take place at agential level—It is discursive and inferential rather than sub-personal and merely intuitive—By rejecting the dissociation of the pragmatic condition of content-interpretation from the semantic condition of interpretation, we should be able to see that it is only with the essential pragmatics that the content can be rendered truth-conditional and thus made 'alive'.

In short, the pragmatics of language-use should be global and primary in terms of explaining an utterance's truth-condition, while rules for 'literal' linguistic contents are rather local and dependent on the truth-condition specified by the context that brings relevant linguistic rules into being authoritative. But the question persists: What, then, is the global pragmatic condition or principle that governs the making of an assertion?

How do we know if a speech act is an act of assertion? This is the demarcation question put in another form.

In the last section, I concluded by suggesting that the Austinian model best illustrates the essential pragmatic condition in assertion. According to Austin, a speech act is inherently an act of illocution. Austin also called the illocutionary act a *performative utterance*. What is peculiar of the performative utterance is that it is at the same time an act expressing a linguistic content and an act interpreting the content thereby expressed as having its meaning saturated. As shown, saying 'you are fired' is an example of performative utterance. There are other examples, e.g.: by making a promise "I will call you tomorrow", I expressed a content with the linguistic meaning of a promise to call you *and* also made that very promise—I made the promise by uttering it in a single token—There is no further act that needs to be done beside my uttering the sentence to make my promise a real one. In short, we can say that a performative act of speech declares itself to be at the same time both authoritative and accountable.[39]

Now, since I argued that we need to regard the pragmatic condition of content-interpretation constitutive of the content's semantic interpretation, we can see that the feature of performatives fits well with our analysis of assertions: In making an assertion "… is true", the agent expresses a 'promise' of the content's truth (later I will call such 'promise' a *discursive commitment*); furthermore, by the same token, the agent's expression is also a description of that 'promise' as being a real one. Therefore, by treating assertion as a performative we can say that an assertion as a speech-act is essentially self-interpreting: namely, it essentially involves the speaker's self-evaluation of the truth of the content as the context of utterance. In later chapters, I will give a more detailed account of the act-theoretic, Austinian assertion. Moreover, I shall also explain how we can evaluate a self-evaluating assertion in a distanced manner, so that an assertion may not be always unqualifiedly true.

[39] Readers might recall the 'dual character' of the laws of logic formulated by Frege (see §1.4.2), namely, the laws of logic in a sense describes the general feature of what is (true) and in another sense prescribes heredity of the meaning of a thought.

CHAPTER 3
The Generality of Force–Content Relation

This chapter discusses the following question: What kind of relation does the relation between force and content exemplify? The question can also be put as follows: When we say that the force demarcates the content and that it is meaning-constitutive, in what exact sense should we understand such kind of relation? How should we understand the relation, in general terms, between asserted and unasserted contents? Answering these questions may help us better understand the essentiality of pragmatics and the idea of pragmatic, discursive agency, which will be introduced in the next chapter.

3.1 Assertion and Content

I suggested in the last chapter that we need to regard assertions as performative speech acts. With this regard, the content's assertoric force can be identified with the speaker's pragmatic, performative force, since the speaker's pragmatic condition of making the assertion is constitutive of the truth-condition of the content asserted. But in doing so, it seems that we are giving an intensional reading of the 'assertoric force'; it then raises the question whether this reading fits Frege's Insight that the assertoric force cannot be assimilated to content while maintaining that the force is meaning-constitutive.

For a theory of formal semantics or semantic minimalism the trick to account for the force while abiding by Frege's Insight is easy: the force cannot be given an intensional reading, otherwise it would be inflationary—Such approach is influenced by Frege's adamant rejection of psychologism; but as I argued, it overlooks the significance of the speaker's pragmatic condition in thinking and in conversing on a truth-conditional content and eventually leads to force–content dualism—The disregard for the meaning-constitutive pragmatic condition has further entrenched the

misunderstanding of the nature of pragmatics, according to which the role that pragmatics plays in our practice of thinking is pushed to two extremes: On one hand the pragmatic effect is thrown to the 'far-side' of the content, signaling the speaker's rather arbitrary enrichment (in Frege's words: 'wording' or 'coloring') of her language, which has no effect on the semantic meaning of her expression's conceptual content; on the other hand, the 'near-side' pragmatic effect is minimalized, so that what is irreducibly pragmatic is attributed to the psychological processes at subpersonal and non-discursive level, e.g. certain 'computational' activities of the mind that are considered to be "no more conscious than the breakdown of sugar in the body".[1] The minimalization of the meaning-constitutive pragmatics leads to the minimalization of semantics, as Pirmin Stekeler-Weithofer pointed out:

> [The fear of (Platonist) metaphysics including all sorts of mentalism has led] to the evacuation of the basic semantic questions into a residual class called 'pragmatics'. Pragmatics now covers the immense realm of questions that deal with all sorts of application of language and theories, including those problems that traditionally belonged to semantics proper. (Manuscript 157)

As I will further argue in this dissertation, the pragmatic condition for a speaker to be engaged in thinking and in conversation is intrinsically agential and discursive. It does not need to be at odds with the content's being truth-conditional; neither does it need to be trivial. For a content to be true is for it to be discursively appropriate—in the sense that it is appropriate to assert the content in the relevant discourse. (And similarly, we can say that for a content to be false is for it to be discursively inappropriate.)

But we must slow down a bit at this point—we want to better understand what does it mean to assert a content: I suggested that an assertion should be viewed as a performative act. Thus, a phrase like 'to make / to perform an assertion' is just a pleonastic way to express 'to assert'. But a phrase like 'to assert a content' is different—It seems to be suggesting that there is an act but also an 'object' to be acted upon—Should the content be seen as an 'object' to be acted upon by the performative forces, we would need to explain how many different kinds of performative forces are there beside the assertoric force and what their relations are to the content.

According to my suggestion, we should take the assertoric force as equivalent to

---

[1] Richard Moran (2001, 6).

the Austinian illocutionary force, and as mentioned, the Austinian illocutionary act is often understood as the speaker's intention. If this is a correct way of understanding, it seems that we can compare an asserted content to an intentional action (an unasserted content, in turn, as an unintended action). But it remains controversial in the theory of action what intentionality means. Consider the famous question raised by Wittgenstein: "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?"[2]—This question seems to imply that to understand intentionality is to find the solution to the equation: *action* + *x* = *intentional action*. Correspondingly, we might also have a similar equation: *content* + *x* = *asserted (forced) content*, and to find out what the assertoric force is we need to subtract the content, say, ⌜*p*⌝, from the whole assertion that *p*. However, it is rather misleading than helpful to formulate the question in this way—As we have seen in §1.4.3, this is exactly the awkwardness of Frege's content-stroke in his *Begriffsschrift*—Since according to Frege's Insight, the force does not contribute to the content, indeed nothing should be left behind if we subtract the content from an assertion. But if we cannot understand the assertoric force by solving the equation above, how should we?

I argued that the correct way to understand the assertoric force—facing the problem of fiction—is to see it as indissociable from a demarcated content regardless of whether it is asserted or unasserted. This coincides with Austin's picture of the 'top-down' pragmatics: According to such a picture, contents are performed in speech acts. Since for Austin the illocutionary force is what a performative force essentially is, and since I take the assertoric force to be equivalent to the illocutionary force, to perform a content is essentially to assert it. Hence, instead of asking what is added to a content to render is as asserted content, we should ask what is added to an asserted content to render it unasserted.

It seems, therefore, that the equation should rather be written as: *asserted content* + *x* = *unasserted content*. But we need to say more to defend this change in the way we write the equation, if we do not want it to be a merely trivial terminological trick. What does it mean, for example, to say that a content is *essentially* asserted? In what sense is the assertoric force indissociable even for an unasserted content? (This is further related to the *question of distanced analyzability* I raised at the end of §1.4.4, which asks how it is possible for us to distance ourselves from an assertion without yet

[2] Wittgenstein (1953, §621).

giving up the condition of discursive intelligibility of that thought. In other words, if a content can only be demarcated by the assertoric force, we still want to better understand the relation between a content's being demarcated as such (i.e. as intelligible content) and its being rendered either appropriate or inappropriate to be asserted.)

While a fuller account of assertion and of the distanced analyzability has to be postponed to later chapters (above all, a detailed treatment of equating assertion with Austinian illocution will be presented in §4.1), in what follows I will focus on how to understand the assertoric force's being essential to the content. I begin by asking the question: If the force is a performative force, namely, if it indicates a type of assertoric speech-act, what kind of generality is exemplified by the assertoric act with regard to its content? To answer the question, I discuss three possible forms of generality: *accidental generality*, *categorial generality*, and *essential generality*. I will show that only in terms of the last form can we appropriately account for the difference of an asserted content from an unasserted one.

3.2 Three Forms of Generality

The idea of distinguishing these three forms derives from Anton Ford's illuminating paper "Action and Generality".[3] As a study of G. E. M. Anscombe's approach of answering the question about the nature of intentionality, Ford asked the same question what form of generality is exemplified in the distinctions between action, intentional action, and mere event. Intuitively, the three candidate forms of generality correspond to three notions in traditional philosophy: accident, category, and essence, respectively. I would like to borrow Ford's threefold distinction for our discussion of assertions, to see which one of the three forms best represents the force–content relation.

3.2.1 Accidental Generality

Describing an assertion in the form of accidental generality involves separating the assertion into two mutually independent features: the content of assertion *plus* the assertoric force. It understands the content and the force as both *sui generis* on their own, and they somehow *intersect* each other to generate an assertion.

[3] Ford (2011).

In a sense, the force–content dualism represents a version of accidental generality: the content is semantically self-contained and autonomous, while the speaker's use-sensitive, pragmatic force is treated as not meaning-constitutive, i.e., as something alien to the content. In short, the dualism divides an assertion into a neutral kernel of content and a pragmatic force, which is usually understood as psychological, attitudinal, or behavioristic elements somehow *added* or *attached* to the content to put the content forward in a token assertoric speech act—In fact, formulating the force–content relation as exemplifying a form of accidental generality necessarily leads to such dualism—We may illustrate the accidentalist model roughly as follows:

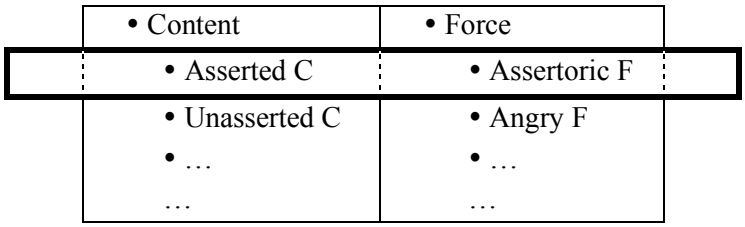

Fig. 3.1

The key to the model is that the two component parts of an assertion are intrinsically independent from each other—The force is not bound to be performed on contents, just like the content is not bound to be forced. There can be other kinds of forces than the force of content (the assertoric force), just like there can be other kinds of contents other than the forced content (the asserted content)[4]—According to this model, an assertion is generated (i.e. the force is assertoric and the content is asserted) if and only if the force and the content 'successfully' intersect each other.

Obviously, the force characterized in this way can only be post-propositional. But since—as I have argued before—the assertoric force must be pre-propositional, the accidentalist model is hence unfit to account for the force–content relation.

Nonetheless, we can say a bit more about accidental generality, to see more clearly why the post-propositional treatment of the meaning-constitutive force does not work. Let us take Aristotle's example of 'round bronze' for example: if there is such a

[4] If we compare the 'force of content' to the propositional attitude, it should be noted that according to the theory of propositional attitude there are more than one kind of attitude towards contents (rather than the assertoric force, there could be different attitudes such as believing, hoping, etc.). But as will be immediately pointed out, such comparison between force and propositional attitude is off-target anyway, since our account of the meaning-constitutive force must be pre-propositional. (On believing cf. §4.3.)

thing, namely a round bronze, then the unity of this thing is explained—*pace* Aristotle's hylomorphic analyses—by the togetherness of round and bronze; but this togetherness needs to be further grounded and thus cannot stand alone as a form of explanation.[5] It is similar in the case of assertions: If we take the category of force to be indicating an agent's various dispositions, then the pragmatically distinctive kind of disposition that a theory of assertion needs is explained by the accidentalist model as exactly the kind of disposition that expresses assertions through its correspondence to the content asserted—Explaining the unity of assertion in this way (i.e. in terms of the 'success' of the intersection of the content and the force as such), however, can at most be trivial, because it does not provide us any substantial criterion for classifying our dispositions (or sentiments, desires, attitudes, etc.) in the way that is competent to single out the valid dispositions for assertions.

The main motivation for researchers to maintain the independency of the force in an act of expression from the content is the conviction that the force, being dispositional or attitudinal, is not truth-evaluable. For example, I can feel angry with another and yet think it makes no sense to do so.[6] Yet this example shows exactly that my angry feeling itself is rather pathological, i.e., it is neither an act nor a judgment. It is not to deny that there must be sensations or qualia that are physically or psychologically primitive for our empirical thinking. But the sensations or qualia as such are not subject to any rational interpretation or evaluation at all. My disposition to be angry, therefore, is not a performative force, neither can my behavior of anger be classified as a kind of thought.

We must distinguish the meaning-constitutive force from forces that are only additive to and hence derivative of the assertoric content. And when we speak of the performative or illocutionary force, we should be speaking of it exclusively as being meaning-constitutive. With this regard, we can see how the so-called 'embedment problem' (also called the 'Frege–Geach problem', as it is originally inspired by the

[5] Aristotle summarizes the feature of the accidental predication as "one thing said of another" (*ἄλλο κατ' ἄλλου λέγεσθαι*) (*Metaphysica* 1030a11), which indicates that the accidental predication is not predication per se, and it always needs to be grounded by some other things. Thus, there must be other forms of predication (such as essential predication) available for us, in order that the subject of an accidental predication could be identified in the first place.

[6] Cf. Gibbard (1990) chap. 7.

Frege Point) can be generally avoided at least at propositional level[7]—The embedment problem refers to the worry that we seem to need indefinitely may kinds of dispositions in our discourse corresponding to indefinitely many different ad hoc semantics of expressive modes in order to account for their embedment in the syntactic structure of our discourse. But we can see that in general there is no anti-luck account of the classification of the 'right kind' of disposition intersecting with the supposedly self-contained truth-conditional contents. As Crispin Wright contends:

> […] we actually lack any clear and workable idea of how to construe discourses which exhibit all the overt syntactic trappings of assertion—negation, the conditional, embedding within propositional attitudes, hypothesis and inference and so on—in such a way that the contents involved are not asserted but are presented with illocutionary force of quite a different kind. (Wright 1992, 11)

The assertoric force is not just one particular kind of forces out of many. Quite the contrary, it is *the* illocutionary force, i.e. the force that renders a content as such. The speaker's other dispositions or attitudes, in turn, rather need to be represented as part of the content.[8] The model of intersection of the force and the content is fundamentally trivial, because we do not understand what a performative force is in the absence of content. In other words, if we say that it is the performative force signified by *S*'s act of assertion that demarcates a truth-conditional content, it seems that we cannot say much about the act other than such (like saying that *S*'s act of assertion exemplifies a specific kind of intention, etc.), because it is the demarcation of the content that enables us to classify *S*'s act, above others, as an act of assertion, not the other way around.

3.2.2 Categorial Generality

Describing the force–content relation in form of accidental generality is depicting them as somehow intersecting. We showed that the model of intersection is, at most, trivial. However, we should not take the force–content distinction to be trivial all together—The question of distanced analyzability indicates that there must be other ways to demarcate a content than asserting it; in other words, there must be an asymmetry

[7] It is another issue that at the inferential level the content of assertion *p* can be recognized as identical with part of the content of the hypothetical judgment 'if P then Q'. See Appendix 2.

[8] Cf. my discussion on the illocutionary points in §4.1.3.

between contents as such on the one hand, and asserted contents on the other.

Understood in this way, it seems that the force and the content, instead of being characterized as mutually independent and accidentally intersecting, should be viewed as rather exemplifying an intra-categorial relation, as illustrated as follows:

- Content
    - Asserted Content (i.e. with Assertoric Force)
    - …

    …

The illustration above removes the entire right column of 'Force' in Fig. 3.1, leaving only the 'Assertoric Force', which is then merged to the left column of 'Content'. According to such model, the assertion is treated as a species of the genus Content.

For the categorial generality, the genus and the species stand in a relation of entailment, e.g., if something is red, it is colored; if something is a horse, it is an animal, etc. Of course, examples of accidental generality can also be seen as instantiating the relation of entailment: something's being a round bronze implies its being a bronze. But such kind of entailment is purely accidental, because being round is not an intrinsic feature of being bronze. Similarly, we can add indefinitely many features to the notion of color at will, say, bright colors, solemn colors, etc. But this cannot be the correct way to understand the generality of the 'species of colors'. In other words, for a categorial generality the taxonomy of a genus should not be trivial. We may group humans into those who drink milk and those who do not; or into those who have two earlobes and those who do not,[9] but the activity of milk drinking and the presence of two earlobes are additive features introduced from outside. The 'differentia' of a species, on the contrary, cannot be assimilated to a predicate.[10]

Ford argues that the differentia of a species is not something independent from that species. He offered the example of colors, and claimed that there is no value that

[9] Though, to be precise, having earlobes is not an accident for humans but a proprium (since only humans have earlobes), but it is nevertheless not an intrinsic quality. We can obviously not refuse to recognize someone as a human if he happens to lack earlobes. Equally, we do not consider a chimpanzee to be human even if it seems to have earlobes.

[10] It is also in this sense that Anscombe rejects the 'predicative view' towards practical reasoning, as if it is theoretical reasoning operating with certain special sort of objects in the practical domain—or like Anscombe ridiculed—as if we can compare the practical syllogism to the 'mince pie syllogism'. See Anscombe (1957, §33)

can solve the equation: *color* + *x* = *red* —For accidental generalities, we can easily clarify the difference of a round bronze from other bronzes in terms of its roundness—In contrast, Ford claims, we would be speechless if faced with the question "What distinguishes red from all other colors?"[11]

Admittedly, we can give a lexical explanation of the species in terms of its genus—What we need may be just some pieces of scientific explanation about the relevant taxonomy: e.g., we may explain that red is the color at the end of the spectrum of visible light that has a predominant light wavelength of roughly 620–740 nanometers, etc. However, as I understand, what Ford claims to be unsolvable in the categorial relation does not consist in the question whether the species could receive any taxonomical definition in terms of its genus. Rather, it consists in that the differentia of a species is not an independently graspable quality but always constitutive of the classification of something under that very species.[12]

We can, of course, define 'red' in terms of 'color' plus 'wavelength'. And scientifically, I see a red object if and only if I perceive light predominantly with such and such wavelength. But such a biconditional-style mutual definition does not help explain how I may classify an object under the species Red in the first place—The description of the wavelength that I predominantly perceive is just another term, and it does not make a definite description after all—In sum, categorial terms are sortal terms, and they should be applicable at the level of singular demonstrations or instantiations, namely, classifications of 'infimae species'.[13] However, the unsolvability of the

---

[11] Ford (2011, 88).

[12] Ford: "No such quality could ever fulfill the explanatory function of a differentia, because it is logically dependent on the explanandum." In idem (2011, 86). If we reduce the differentia of a species to a predicate, it would be like an extra feature of an object that we can recognize independently of the discernment of its color—It would lead to the view of 'highest common factor' in epistemology, according to which the genus can be seen as having an independently recognizable common factor shared by all its species, so that we may, for example, see a colored object without sensing its specific color until later when the wavelength of the light is determined in a fashion of enhanced accuracy. However, this cannot be the way how we single out an object as falling under a category. (We can, of course, first see a car and then later discern that it is a Volkswagen, but for this reason the Volkswagen is not, strictly speaking, a species of cars.) A critique of the common factor view cf. McDowell (2006).

[13] It has been argued, for example, that since the notion 'species' has biological implications, we should differentiate it from natural kinds, so that species are individuals rather than classes. Cf. Hull (1978). I will remain neutral on this issue, yet it suffices for me to show that the difficulty

equation of differentia indicates that—just as shown in §2.3—the singular instantiations of empirical concepts are always context-sensitive, and introducing further predicates does not help eliminating the context-sensitivity. (Enumerating different species of colors, for example, is notoriously an open question, because there seems to be indefinitely many of them.)

Now, it might seem that the feature of categorial generality is adequate to account for the generality of assertions: the assertoric force is a differentia that distinguishes an asserted content from an unasserted one, and it coincides with Frege's Insight that the force, which is the differentia, cannot be assimilated to content. And similarly, there is no comprehensible solution to the equation: *content* + *x* = *assertion*. Nevertheless, I think the categorial generality is not a satisfactory candidate. According to the categorial model, although the force does not add anything to a content like a normal predicate does, it nevertheless classifies or categorizes the content as an assertion. But it falls short of explaining how a non-predicative classification should be understood.

This is a problem not just for assertions, but also for all kinds of regular categories. To differentiate categorial from accidental generality, we need to be able to demonstrate that it is intrinsic rather than accidental to something's property of being red that it is colored. But by maintaining the unsolvability of differentia, it remains unclear how the standard for classifying something as red entails in a non-contingent way that it is also colored. As I will indicate in the following subsection, in order to really establish the assertion–content relation as non-accidental we need *per se* predication, which, in fact, exemplifies a form of essential generality.

3.2.3 Essential Generality

We have seen in the last subsection that the description of the feature of categorial generality is merely negative. In fact, this is the aporia of intra-categorial predication that Aristotle has shown in his *Metaphysics*. He argued that the unsolvability of the equation of differentia actually consists in that the predication of a differentia is neither intrinsic to the predication of its genus nor to the predication of the species.[14] The

---

in accounting for categorial generalities is immense, and it seems unlikely to be surmountable just with the descriptions of the categorial generality that we have here.

[14] See Aristotle *Metaphysica* 998b17-28, also *Topica* 144b4-9.

argument about categories is very complex, and it is certainly beyond the scope of this dissertation to rehearse or discuss the argument here. Instead, I would like to simply jump to Aristotle's positive conclusion: There must be certain *primary species* for a genus, namely, primary being that involves a form of *per se* predication[15]—Unlike that if *A* is a round bronze, then *A* is round *per shape*; we instead say that if *A* is a human, then *A* is rational *per se.*[16] Moreover, for a primary species, Aristotle claims, the predication of a genus and the predication of a differentia would be identical—It means that one's being rational is the essential quality for her to be human and vice versa; and in this case, we are speaking of her *essential generality*.[17]

Ford explains the feature of essential generality in that it can be exemplified by the equation: *gold + x = pure gold*. Like the categorial generality, there is no comprehensible solution to such an equation. But this time, the equation also means that there is nothing said of pure gold beyond the notion of gold. The 'purity' neither introduces any extra feature to the genus nor classifies it, but is just an indication of the perfect, *unqualified constitution* of being gold. Similarly, we can label 'loyal friend' and 'perfect circle' as the primary species of Friend and Circle, respectively, because a friend without qualification must be loyal, and a circle without qualification must be perfect.

Although the genus Color has indefinitely many species, it does not have a paradigm color. By contrast, under the formula of essential generality, even though there could similarly be indefinitely many species of gold: gold that is 75% pure, that is 90% pure, that is 99.95% pure, and so on, these varieties have nothing comparable to

---

[15] Cf. e.g. Michail Peramatzis (2010).

[16] For Aristotle, this is the contrast between *kata symbebekos* (*κατὰ συμβεβηκός*) and *kath' hauto* (*καθ' αὑτό*).

[17] According to Aristotle, the category of a thing designates its being *per se*, i.e. it designates what it is by virtue of itself. It is thus somehow ironic that the ordinary categorial model of generality fails to really capture the category of a being. Kosman has in his (2013) pointed out that Aristotle's essentialism is not a theory of predicative property but of the constitution of subject:

> Aristotle's essentialism, in contrast to Aristotelian-style Essentialism, is not primarily a theory concerning properties at all, as essence is not so much for Aristotle a matter of predicates as of subjects. It is a matter not of a predicate being said one way or another of a subject, but of the necessary constitution of subjects, a constitution that is founded on the unique suitability of substance to serve as ultimate subject. (173)

the notion of the pure gold itself; Moreover, they are understood only in relation to the idea of pure gold—The pure gold is the central case of being gold—In Aristotelian terms, it is the principle (*arché*) for the identification of the indefinitely many non-central and hence derivative cases.

While the 'traditional' philosophical notions such as essence or arché might seem difficult, we can compare them in our context to the problem of meaning: If we regard the genus as an equivocal that denotes a range of possible meanings, the primary species, then, denotes its explanatorily primary, default meaning (which is also called 'focal meaning'). It unifies the multiplicity of possible meanings of the equivocal into an equivocal *pros hen* (equivocal with focal meaning). Therefore, it is not the genus (e.g. the literal, equivocal term) that unifies its species, rather it is the primary species (the focal meaning of the term) that sustains the unity.

For this reason, strictly speaking it would be improper to say that the primary species 'falls under' the genus. Pure gold is not *a kind of* gold. It is not just one out of many; rather, it *is* gold, without qualification—It signifies nothing more than the absence of impurities—Correspondingly, the differentia of a non-central species is its qualification or privation with respect to the primary species. I may encounter indefinitely many varieties of gold in reality, but they are associated as varieties *of (impure) gold* only by being dependent on the primary species of pure gold.

Now, let us consider assertions. Formulating the generality of assertion in terms of essential generality means that assertion is the primary species of content, i.e., it is content without qualification, while other unasserted sorts of sentences and expressions (fictions, indirect speeches, etc.), by contrast, must be understood as derivative and thus qualified with respect to assertions—as far as they are to be treated as contents at all. In this sense, we can thus rather write the equation like: *asserted content* + *x* = *unasserted content*, where the '*x*' indicates the qualification (i.e. the distancing) for an unasserted expression to be understood as truth-conditional content.

This form of generality of assertions coincides with the Austinian, 'top-down' pragmatic account I proposed in Chapter 2. The 'top-down' pragmatic account reverses the customary 'bottom-up' view that contents have primacy over assertions in representing its truth-conditions. Instead, it should rather be the performative act of assertion that is explanatorily primary in representing a content's truth-condition. Recently, Peter Hanks has argued for the similar idea in his book:

> [According to a misleading Fregean picture,] propositions are out there, with their truth conditions intact, waiting to be judged and asserted. A subject latches onto one of these propositions—Frege calls this 'grasping a thought'— and then goes on to endorse it in thought (judgment) and put it forward as true in speech (assertion). […] I believe we need to reverse the explanatory order. Propositions get their truth conditions from particular acts of judgment and assertion, which are themselves the original or primary bearers of truth and falsity. The source of truth conditions is to be found in the acts of representation we perform when we make judgments and assertions, not in the propositional contents we use to classify and individuate these actions." (Hanks 2015, 3f.)

This idea hence puts forward a generally 'act-theoretic' approach to contents. According to the act-theoretic view, to think is to make a move in the language game, namely to represent a content in the logical space of reasons concerning its truth-conditions, which can be seen as a practice of discursive engagement. Assertion, then, is the paradigmatic form of our discursive engagement, which expresses that the content performed is discursively appropriate. The unasserted expressions such as fictions, hypotheses, imperatives, and interrogations are nonetheless discursive acts, i.e., acts conversing on truth-conditional contents, although they are non-central and hence derivative of the performative act of assertion.[18]

Viewing assertion as exemplifying essential generality also means that assertions, strictly speaking, are not *a kind of* content (or a kind of speech act). In other words, the assertion is not one mode of presentation out of many. Rather, it is presentation of a truth-conditional content *qua* truth-conditional content.[19] The assertion and the distancing (which Hanks calls 'cancellation'), therefore, are neither two separate types of acts, nor to be unified under any higher-order 'super-type' of act; rather, they should be seen as an asymmetrical pair of notions that refer to a singular type of act, i.e. the discursive act, where the assertion denotes its central case.

---

[18] Hanks claims that interrogations and imperatives cannot be subsumed under the form of assertion. Though I cannot discuss this in detail, I believe our analysis is capable of explaining that there is no reason to regard interrogations and imperatives as incompatible to the general framework of the discourse of commitment to the truths (appropriateness) of contents or its inhibition that allows others to ask Whys and Hows. See Appendix 3.

[19] This may be reminiscent of Davidson's famous claim about intentional action: "intentional actions are not a class of actions, or, to put the point a little differently, doing something intentionally is not a manner of doing it." See Davidson (1967, 121).

In summary, we can regard assertions as exemplifying a form of essential generality relating to contents. The accidental generality is inadequate in this regard because it conflicts with Frege's Insight and treats the assertoric force as an additional predicate. The categorial generality, on the other hand, coincides with Frege's non-predicative view of the force; but it nonetheless fails to provide any positive account of the force while it still considers the force as content-classifying—E.g. it is incapable of explaining how the assertion could be unique among all sorts of contents rather than being a regular kind of content (as if it is like a kind of color in an open-ended list of different species of colors). The form of essential generality, by contrast, does not render the assertion as just a kind of (presentation of) content; rather, the assertion is viewed as unqualified content—If we could still maintain that the assertion–content relation exemplifies a form of species–genus relation at all, we must comprehend assertions as the primary species for all contents. In this regard, the assertoric force, i.e., the differentia, neither introduces anything extra to the content, nor functions as a regular classifier—In unqualified cases, the force indicates illocutionary performance without qualification, while in qualified cases, the force is distanced, which indicates qualifications of illocutionary performance. In the next section, I would like to suggest that we can regard the force as signifying the *constitutive unity* of the content.

3.3 The Constitutive Unity of Contents

Adopting the form of essential generality for assertion–content relation means that a content is essentially assertoric, i.e. essentially forced. In other words, it means that the assertoric force is constitutive of any content. In §3.3.1, I would like to examine what it means to say that one thing is constitutive of another. Based on this, I will discuss in §3.3.2 my proposal that the force signifies the constitutive unity of content.

3.3.1 Relations of Constitutivity

In this subsection, I shall distinguish two ways to understand what the phrase "being constitutive of" means. The *first way* is to give it a conditional reading—It understands "*A* is constitutive of *B*" as expressing a form of material implication—In this sense, we can understand the relation of constitutivity as a binary function, which can be written as follows:

(1) $\lambda x\, \lambda \boldsymbol{X}\, \lambda \boldsymbol{Y}$ [(CONSTITUTIVE OF) ($\boldsymbol{X}(x)$, $\boldsymbol{Y}(x)$)] = $\lambda x\, \lambda \boldsymbol{X}\, \lambda \boldsymbol{Y}$ [~(~$\boldsymbol{X}(x)$ & $\boldsymbol{Y}(x)$)]

This way of understanding constitutivity can be used to express the relation of entailment between species and genera as in the form of categorial generality. It says, therefore, that something's belonging to a corresponding genus is constitutive of its being a species. For example, since something's being colored is implied by its being red, we can say that it is constitutive of its being red that it is being colored, so that it would be wrong to maintain that it is not colored but red. Let $\boldsymbol{X}$ indicate the genus Color and $\boldsymbol{Y}$ the species Red, we have:

(2) $\lambda x\, \lambda \boldsymbol{X}\, \lambda \boldsymbol{Y}$ [(CONSTITUTIVE OF) ($\boldsymbol{X}(x)$, $\boldsymbol{Y}(x)$) & (COLORED) ($\boldsymbol{X}$) & (RED) ($\boldsymbol{Y}$)]
= λx [~(~(COLORED) (x) & (RED) (x))]

However, obviously we cannot apply such reading of constitutivity to the force–content relation. It would need to treat the content and the force as two separable predicates. But since the force is non-predicative, it does not make sense to write something like 'λx [(FORCED) (x)]'; on the other hand, the formula 'λx [(CONTENT) (x)]' would be as superfluous as Frege's content-stroke.

In fact, treating the relation of constitutivity as expressing a form of material implication does not square with the model of essential generality in general. When applying material implications to categorial relations, we said that something's genus is implied by its species. But in cases of essential generality, it is actually trivial to say that something's genus is implied by its being a primary species of that genus. Therefore, we do not say that being gold is constitutive of being pure gold; Instead, we say that being pure (gold) is constitutive of being gold.[20] Likewise, we say that being forced is constitutive of being content, rather than the other way around, and so it does not express any relation of one thing entailing another.

Therefore, in this *second way* to understand constitutivity, '*A* is constitutive of *B*' does not simply express any conditional relation between the two—For *B*, *A* is not simply expressed as a necessary and/or a sufficient condition as such. Rather, the

[20] Just like the assertoric force, here being pure should not be seen as a predicate dissociable form gold—We may regard it as a non-predicative notion reserved for signifying the unqualified constitution of gold—We may, for example, call the purity of gold 'gold purity', or call the perfection of circles 'circular perfection', etc. The wording here is admittedly awkward, but this just shows that the term for the unqualified constitution of a type is actually not a dissociable predicate.

constitutivity is more than that: '*A* is constitutive of *B*' says that it is in terms of *A* that a *B* could be comprehended as *B* in the first place. In other words, if *A* is constitutive of *B*, comprehending *B* in terms of *A* is explanatorily primary for any comprehension of *B*. Comprehending *B* in terms of *A*, in other words, signifies the primary, paradigmatic species of *B* (i.e. *B*'s focal meaning). Therefore, in this second way of understanding constitutivity, being forced is constitutive of being content, since, as argued, forced content is the primary species of content. (And we can also say that being pure is constitutive of being gold, and being organismal is constitutive of being healthy, to the extent that pure gold is the focal meaning of gold, and organism health is the focal meaning of health, etc.)

Furthermore, only in the second way of understanding constitutivity, if *A* is constitutive of *B*, *A* is also the 'demarcator' of *B*—In the first sense of constitutivity, 'being colored' does not demarcate a red color from other colors. But in the second sense of constitutivity, we can say that (gold) purity demarcates gold as gold, or (circular) perfection demarcates circle as circle, etc., just like that the force, as argued, demarcates a content as content. I suggest that in this second sense *A* can be actually seen as the *constitutive unity* of *B*—as I will argue in the next subsection.

3.3.2 Constitutive Unity

The assertoric force demarcates a content from other occurrences of our mental life. A forced content is a meaningful content, the primary species of all contents. The force thus signifies the constitution of conceptual content proper, just like that the gold purity signifies the constitution of gold proper. But not all demarcators signify the paradigmatic constitution of their 'demarcatees' in this manner. For example, in ordinary arithmetic, the number 0 is demarcated by the Peano axioms from other numbers. But such demarcation of the number 0 is purely analytic under the Peano axioms—The demarcator and the demarcatee, in other words, do not exemplify a relation of essential generality (They are rather tautological)—By contrast, when the assertoric force demarcates a content, it not only demarcates what it is, i.e. a content, but by being constitutive of the content it also signifies the content's paradigmatic

constitution. Generally, if a demarcator signifies the paradigmatic constitution of its demarcatee, we can say that it signifies the *constitutive unity* of the demarcatee.[21]

The force signifies the constitution of content proper. But a constitution may be either proper or improper, just like a content might be asserted or unasserted. It seems that we can compare a content's constitution to its truth-condition—For this reason, we can see that there is a close relation between a content's propriety (its being constituted as content proper) and its being asserted as true—As Dummett states:

> […] our concept of truth gets a large part of its point from the contrast that we wish to draw between a statement's being true and any more primitive, or at least undifferentiated, conception of its being appropriate: for instance, between its being true and the speaker's having a sufficient warrant to take it as true, or between its being true and the intention that the speaker had in asserting it to be true just then being a just one, his having had a legitimate point in making it. (Dummett 1979, 126)

Indeed, as I have earlier indicated, I take the truth of a content to be synonymous with its discursive appropriateness. To assert a content, therefore, is equivalent to taking an expression as an appropriately constituted content in discourse.

Since the force not only demarcates a content as content but also signifies the propriety of the content's constitution, we can regard the force as invoking standards or norms for evaluating the content's truth—This is also the case for gold: When we say that the gold purity signifies the propriety of the constitution of gold, the notion of gold purity implies a series of physical and chemical standards for demarcating gold—However, what is special of the assertoric force is that it not only implies norms for content-evaluation, these norms are essentially pragmatic, i.e. they are also norms constitutive of the performance of the content in our discursive practice of thinking and speaking. In the next chapter, I shall take a closer look at the significance of the pragmatic, performative unity of content, which I will discuss through the notion of the constitution of discursive agency. It will allow us to inquire more about the assertoric

[21] The notion 'constitutive unity' has been used, for example, by Andrea Kern, who explicitly distinguishes the constitutive unity from the analytic unity. Cf.: "[…] we can call the sort of unity exhibited by a rational capacity a non-analytic or, positively, a constitutive unity. By a 'constitutive unity' I mean the sort of unity whose elements are characterized by their logical dependence on the unity they jointly make up." (Kern 2017, 148) Readers might also recall Kant's distinction between the synthetic and the analytic unity.

force and the constitutive features of contents, such as commitment, belief, and first-person thought.

CHAPTER 4
Discursive Agency

In the previous chapters, I argued that we should adopt a broadly Austinian, 'top-down' view of the pragmatic condition of assertion. According to this view, truth-conditional contents are inherently assertoric. It thus indicates a speech-act-theoretic approach to contents and the assertoric force—I suggested that the assertoric force should be understood as a pragmatic, performative force of utterance, which signifies constitutively the unity of the content. In other words, the unity of a conceptual content essentially consists in the unifying, performative act of assertion, so that we can equally speak of a pragmatic unity of assertion or a pragmatic unity of thinking in general. In this chapter, I will lay out some fundamental arguments about the pragmatic unity of thinking through the notion of agency.

To do this, I shall begin in §4.1 with a detailed introduction to the Austinian treatment of speech acts and discuss in which sense we can bring the broadly Austinian view and the broadly Fregean view together. The notion of constitutive, discursive agency will then be examined in §4.2 according to the act-theoretic, commitment-based approach to assertions. I will discuss how we should account for beliefs and first-person thoughts under this approach in §4.3.

4.1 Assertion and Illocution

4.1.1 Illocutionary vs. Locutionary Acts

The approach I will defend can be seen as an 'act-theoretic' approach, since it attaches much importance to the pragmatic aspect of the utterer as an agent of performative utterances, which reminds us immediately of the speech-act theory. The speech-act theory has attained remarkable development and influence since the middle of last century thanks to Austin's significant work. And particularly, there has been much attention drawn in recent years to the approaches to the propositional contents and its

relating Fregean problems based on a generally act-theoretic framework.[1] In general, according to the act-theoretic approaches the demarcation of a thought asks for the constitutive demarcation of the performance of uttering that thought, while such performance is regarded as a *sui generis* type of cognitive act.

The act-theoretic approach thus explicitly combines the Austinian idea of the performative utterance in the speech-act theory with the Fregean idea of the assertoric force—Considering our previous discussions on Frege, we can see that this is a natural outcome, since the Austinian speech-act theory and the Fregean project of the Concept Script, in a sense, are working on different sides of a same issue: The Fregean logical project starts by asking what describes the essence of thinkable contents as demarcated from other intrinsically unintelligible contents and concludes in the primacy of the content of *assertion*, i.e. the content that has a unity of meaning by virtue of its constitutive assertoric force; By contrast, the Austinian speech-act theory starts by asking what describes the essence of an act of speech as demarcated from other kinds of acts and concludes in the primacy of the act of *illocution*, i.e. the performative act of conveying meaning in using language. Hence, from this perspective, both the Fregean assertion and the Austinian illocution intend to establish a unity of use and meaning of contents.

For Austin, what is of most importance for the *illocutionary acts* is that it should be differentiated from *locutionary acts*. In his lectures published under the title *How to Do Things with Words*, he distinguishes three kinds of locutionary acts:

> [We can make] three rough distinctions between the *phonetic* act, the *phatic* act, and *rhetic* act. The phonetic act is merely the act of uttering certain noises. The phatic act is the uttering of certain vocables or words, i.e. noises of certain types, belonging to and as belonging to, a certain vocabulary, conforming to and as conforming to a certain grammar. The rhetic act is the performance of an act of using those vocables with a certain more-or-less definite sense and reference. Thus 'He said "The cat is on the mat"', reports a phatic act, whereas 'He said that the cat was on the mat' reports a rhetic act. (Austin 1975, p.95, my emphases)

[1] Most of these recent approaches, such as Hanks (2015), though, do not specifically connect themselves with the Austinian speech-act theory. Cf. also Recanati (forthcoming), and Scott Soames (2010), (2014), (2015), (forthcoming).

Both the phonetic act and the phatic act are incomplete for obtaining meaning by themselves. While a sheer phonetic act utters only noises, a sheer phatic act only 'quotes' or 'mentions' certain expression in a language without being considered as really comprehending and using that language—As discussed at the very beginning of this dissertation, the expression expressed by an act as such (like what a computer could do) does not inherently convey meaning—Although a phatic act may express a full, grammatically correct sentence, the sentence itself remains 'quoted' and has no semantic value on its own.[2]

The rhetic act, interestingly, expresses sentences that are disquoted and therefore already have certain semantic value. But it is nevertheless regarded by Austin as merely locutionary because he who said that the cat was on the mat was reported only through indirect speech (*oratio obliqua*). It means that it is not the reportee's act of saying so-and-so that demarcates the expression as meaningful. The content of "so-and-so" may have semantic values, but it has them only derivatively, because it is *only* disquoted, demarcated as thought, and hence qualified as having meanings *when* it is reported in an indirect speech. In other words, it is primarily the *reporter*'s act—namely, the act of direct speech (*oratio recta*)—of reporting that the reportee said so-and-so that it bears a pragmatic force and thereby makes a move in the language game. Therefore, Austin concludes that "To perform a locutionary act is in general, we may say, also and *eo ipso* to perform an illocutionary act." (Austin 1975, 98).

4.1.2 Illocution vs. Intention

At this point, we can draw a comparison between illocutionary acts and intentional actions: The recognition of a locutionary act as a performative speech act requires the recognition of an explanatorily primary illocutionary act, just like that the recognition of

[2] Recall John Searle's well-known example: Suppose an American soldier was captured by several monolingual Italian soldiers in the Second World War and tried to convince his captors that he is a German officer. He recited the only line of German he could remember, which he learned during his school days: "Kennst du das Land, wo die Zitronen blühen?"—He intended his hearers to believe that he was saying, in German, that he is a German soldier, and he succeeded in doing so and eventually got himself released. But in this case, the sentence uttered by the American soldier (and interpreted by the Italian soldiers) only performed the function of phonetic imitation. What the American soldier actually intended to achieve, and what the Italian soldiers thought he meant to say, was not expressed *in* the utterance of the German line as such. See Searle (1969, 44f.).

a bodily movement requires the recognition of an explanatorily primary intentional action. It is noteworthy, however, that although such a comparison between illocution and intention is helpful, the two are not simply identical.

We need to see that speech act theory is more general than practical deliberations and actions: The performance of illocutionary acts has yet nothing to do with the distinction between theoretical and practical knowledge and with the debates on the directions of fit—Austin thus differentiates the illocutionary act from another derivative kind, namely the perlocutionary act, which may serve as the transition from the general pragmatics to the specific domain of practical knowledge, e.g. deliberation of the practical consequence of the utterance "I persuaded him to shoot her". I will not discuss perlocutionary acts in this dissertation, since the distinction of directions of fit is not central to our topic. I presume that both the theoretical and the practical knowledge, as far as they convey meaning and express knowledge at all, require a basic unifying account of the performative force of thinking and assertion.[3]

Admittedly, many philosophers do talk about the speaker's intention in making an utterance—A remarkable example is Grice's use of the term 'M-intention': By

---

[3] Cf. Austin: "the performance of an illocutionary act is a "performance of an act *in* saying something as opposed to performance of an act *of* saying something." (Austin 1975, 99). It indicates that even though assertions or illocutions have certain aims, e.g. knowledge, conviction, etc., they are not specific consequences which assertions and illocutions serve as instruments to achieve.

In his book, Hanks wants us to put the practical aspect of an assertion to the background and foreground the theoretical aspect—His example is the contrast in statements like "What he said is true, but what he did is bad" (Cf. Hanks 2015, 70f.). But we also need to see that it is not only out of our arbitrary interest to foreground the theoretical aspect of the speech; rather, it is because the theoretical–practical division is secondary to the question of illocutionary, speech-act agency.

Moreover, there is in fact an increasing trend in recent studies to argue that practical intentions and voluntary actions share a same logical form as (theoretical) judging and believing, which largely blurs the line of the theoretical–practical division. Such a view usually originates from the Anscombean approach, which argues that the practical 'doing' *is* thinking. I am also sympathetic to this view. For discussions see my (manuscript). Cf. also McDowell (2011), Rödl (2007) chap.2, and idem (2010), Thompson (2008), (2011). An important conclusion of this view is that all our (non-mathematical) judgments must be constitutively 'practical' if there is such a thing as 'practical judgment'. This is not to say, however, that we judge and believe as if we could create truths at our arbitrary will; quite the contrary, it shows that it is problematic to claim even for voluntary actions that we could create truths arbitrarily—in fact, our practical judgments and actions are fallible just as the theoretical judgments do.

uttering φ, the speaker *S* M-intends that *p* iff *S* means that *p* by uttering φ—Such terminology is fine, as long as we bear in mind that in taking the illocutionary act to be intentional it should not be confused with practical intentions. Rather, it should be seen as a sui generis kind of intention (i.e., the intention of conveying meanings through discursively appropriate conceptual contents). In this sense, the illocutionary act is constitutive of and much more fundamental than any other kinds of intentional acts—As I shall later argue, it is not a regular intention that an agent can decide to have or not to have. It is not an 'optional' intention that is additive to the content; instead, it is constitutive.

Comprehending the illocution as a more general and sui generis kind of intention helps us better understand the distinction between illocutionary and locutionary act. Searle once criticized that Austin failed to make this distinction. Consider the following example:

(1) A told B that the cat is on the mat.

(2) A informed (or notified, etc.) B that the cat is on the mat.

According to Searle's Austin, while Sentence (1) describes a locutionary (i.e., rhetic) act, Sentence (2) describes an illocutionary act.[4] Searle further contended that this distinction is problematic, since both (1) and (2) should be understood as illocutionary act. In a sense, Searle's criticism is not unjustified, because Austin did use similar examples in his lectures for the locution–illocution distinction and he was indeed not clear enough in giving those examples involving speech-act verbs. But Searle's criticism was nevertheless based on a wrong interpretation of Austin. For Searle's Austin, what distinguishes the verb 'to tell' from 'to inform' is that the former is a neutral act of saying so-and-so, while the latter, by contrast, describes the speaker's certain specific intention in the act of saying so-and-so (e.g. the intention to inform). However, we should see that for Austin the issue here is not whether this or that verb describes a specific intention of the speaker; rather, what really motivates the distinction between the locutionary and the illocutionary act is the direct–indirect speech distinction. Therefore, according to a charitable reading of Austin, if the verb 'told' in Sentence (1) describes a locutionary act, it is rather because it invokes the content in the that-clause as a report—namely, as an indirect speech—embedded in a direct speech.

[4] This was not exactly Searle's original example. I have modified his example a bit.

Equally, when Austin said that the verbs 'to inform' or 'to notify' describe an illocutionary act, it is not because these verbs represent specific intentions *beside* saying so-and-so, but because of the general intention to convey meaning—the intention that is constitutive of any act of speech.

Instead of 'intention', I think we should better have a separate term to describe this special pragmatic character of the illocutionary act—I think 'commitment' is an appropriate term for this. In §4.2, I will shed some more light on the notion of (illocutionary) commitment—or to be more accurate: discursive commitment—and discuss how it could help us establish a feasible act-theoretic approach. But before that, I shall first make some clarifications about the conventional treatment of illocution and assertion in the speech-act theory and how it can be modified into our broadly Fregean approach to the assertoric force.

4.1.3 Assertion vs. Illocutionary Points

By distinguishing the illocutionary act from the locutionary, Austin summarized that the positive understanding of the illocutionary act is that it has "certain *force* in saying something." (120) He then went on to explain how this force should be understood. (In a sense, his whole lectures could be seen as an attempt to distinguish and describe the specificity of the force of illocution.) Austin's detailed treatment of the illocutionary force aside, I suggest that it provides a good perspective for us to understand Frege's assertoric force. In fact, I would like to claim that they are the same kind of force, and the act of assertion and the act of illocution are the same kind of act.

Frege's undertaking of the assertoric force obviously influenced Austin to use 'force' to explain illocutionary act and its relation to meaning. However, there seem to be an immediate discrepancy between the speech-act theory and Frege's Concept Script: As discussed, in Frege's Concept Script there is one and only one kind of force, namely the assertoric force. But according to the speech-act theory there are multiple subspecies of the illocutionary force—Searle called the multiple subspecies the *illocutionary points*, and summarized a total of five different illocutionary points in the English language: Assertive, Commissive, Declarative, Directive, and Expressive.[5] It thus seems that the assertoric force only corresponds to one of the many illocutionary points,

[5] See Searle and Vanderveken (1985, 13f.).

namely the assertive, so that illocutions cover a larger scope than assertions. If so, it would be obviously at odds with my proposal of equating assertion with illocution.

When Searle introduced the illocutionary point, he wanted to classify the different kinds of intentions a speaker may have in her performative utterance. But when it comes to classifying the speaker's intentions it is always a question how fine-grained those distinctions should be and whether the list could be complete—If the principle for the classification stems from descriptions of linguistic phenomena (e.g. descriptions of the variety of grammatical moods of speech), we obviously must expect the taxonomy to be highly indefinite—The question concerns not only Searle's illocutionary points but also other various linguistics-based approaches to the taxonomy of illocutionary acts.

Though a linguistics-based approach could be significant in its own right, our interest is instead to account for the basic philosophical features (of general demarcation, of truth and meaning, etc.) of the illocutionary act as such. Therefore, when I speak of assertions and assertoric force, I always have Frege's Original Insight in mind: the assertoric force must be meaning-constitutive, and it cannot be assimilated to a predicate. As the only performative sign in his Concept Script, the assertion sign is not a mood indicator; rather, it serves as the demarcator of meanings. The assertion is thus synonymous with judgment, which is neither a regular mode of speech nor a post-propositional add-on of far-side pragmatic effects. For this consideration, I think it is appropriate to adopt a broad-brush approach of equating assertion with illocution, because they both signify the constitutive unity of the content.

One might object at this point that the seeming departure of my approach from the customary approaches of speech-act theory is only terminological: namely, I call the illocutionary act 'assertoric act' only arbitrarily, while my interpretation of assertion has nothing to do with the assertive mode of speech in the speech-act theory. However, on the one hand, I do not think it is purely accidental that Frege uses 'assertion' to describe the act of judgment and affirmation; and on the other hand, even in linguistics-based speech-act researches the assertive mood as a mode of speech also enjoys certain peculiarity among other grammatical moods—namely, the assertion is illocutionarily basic, because asserting that *p* is nothing but presenting *p* as a conceptual content by meaning it (i.e. taking it as true). Consider the following example:

(3) The cat is on the mat.

(4) I assert that the cat is on the mat.

The gist of assertion—for both the Fregean project and the speech-act theory—is that Sentence (4) says the same as Sentence (3), namely, both have the same truth-condition. In other words, the act of asserting adds nothing more to the content asserted.

Sentence (4) says nothing more than that which the content in Sentence (3) says, because in mentioning the act of asserting it does not add any mood to the content, nor does it report an indirect speech, but signifies a direct speech. Therefore, as William Alston correctly puts: "To assert that *p* is to R that *p* and to do nothing more in the illocutionary line."[6] Alston uses the letter 'R' to indicate verbs that can describe a speech act. And in the case of assertion, we can use many different verbs in English to achieve the similar effect: instead of using 'I assert', we can also use 'I affirm', 'I declare', 'I claim', 'I announce', 'I agree', 'I conclude', … and so on. These verbs may have further different conversational implicatures on their own, but those differences do not concern the general constitutive pragmatic condition of asserting the content.

In his book, Alston summarizes as many as 29 speech-act verbs that fall under the broad category of assertion.[7] But as I claimed, the issue about assertion here is fundamentally an issue not of linguistic categorizations, but of the nature of thought-presentation and its theoretical paradigms. Therefore, if we adopt the broad-brush approach of equating assertion with illocution, we make a claim stronger than those of the linguistics-based speech-act theories: The stronger claim is that assertion is not only linguistically peculiar but also fundamentally different from other moods because it indicates the force constitutive of any content with any possible moods. With this perspective, we may call the central, paradigmatic act of presenting a thought the act of illocution, or judgment, or assertion, or—as later will be shown—discursive commitment, and they are all the same. In a sense, it is indeed a terminological issue, but it is not *only* terminological. What is important is the philosophical understanding behind these terms.

---

[6] See Alston (2000, 114).

[7] See ibd. p. 125.

4.1.4 Asserting vs. Distancing

There is yet another, much more serious objection to my equation of assertion with illocution: According to the customary speech-act theory, although there are many speech-act verbs that are illocutionary *and* assertive—such as those I mentioned earlier—there are nevertheless speech-acts verbs that are illocutionary but *not* assertive, such as command, promise, and suggest, etc.

To reply, I think we can distinguish two senses of assertion: a narrow and a broad one, respectively. In the narrow sense, the assertion corresponds tightly to a mode of speech. And in this sense we can clearly say that some other speech modes, say, the directive, is not an assertion. (We might want to use the term 'assertive' specifically in this narrow sense.) Assertion in the broad sense, however, means the same as illocution, namely, presenting *p* by meaning it. And in this sense we can say the directive is also an assertion.[8] In this dissertation, I am exclusively interested in the assertion in the latter, broad sense.

Of course, not every sentence is an assertion. But according to my broad-brush approach of equating assertion with illocution, saying that not every sentence is an assertion means just the same as saying that not every locutionary act corresponds to an illocutionary act. However, we also indicated that it is central to Austin's undertaking that the illocutionary act is constitutive of and explanatorily prior to the locutionary act. In fact, equating assertion with illocution can help us better understand the question of distanced analyzability and respond to the Frege Point—In the previous chapters I argued that the Frege Point which states that a same content can be once asserted and once unasserted with its truth-condition remaining intact is problematic. The Frege Point entails a correct observation that even an unasserted content, if it is demarcated as conceptual content at all, should be possible to be identified as preserving the same content as a certain asserted one.[9] But Frege drew the wrong conclusion from this observation that the unasserted content is therefore a kind of self-sufficient proposition free of any force, and the assertoric force is hence just a post-propositional add-on.

---

[8] Though I cannot discuss this issue in detail in my dissertation, I will give a brief remark on it in Appendix 3. In short, I do not think that some modes of speech, such as interrogatives and imperatives, are categorically different from that of assertions and hence cannot be brought under the broad-brush approach that I am suggesting here.

[9] Discussions on content-preservation cf. Chapter 5, 6 and Appendix 2.

Contrary to this, I suggested that we should insist on the consequence of Frege's Original Insight, namely that the assertoric force is constitutive of any thinkable content, so that the force is rather pre-propositional—The critical idea to explain the observation is that the unasserted content is not *less than* but rather *more than* a default asserted content—As Dummett correctly puts:

> [I]f we used an assertion-sign, the actor would have to use it too. The reason he is not making assertions is not that he is doing *less* than that—merely expressing thoughts, say—but that he is doing *more* than that—he is acting the making of assertions. What constitutes his doing this is his uttering the assertoric sentence—with the assertion-sign if we have one—in a context […] on the stage in a theatre at an announced time. (Dummett 1981, 311)

An unasserted content cannot be seen as independent of the assertoric force while remaining nonetheless thinkable. Though we may regard the unasserted content as 'non-forced', it is not really 'bare' or 'force-free'. It rather means that its force is distanced or cancelled, while to distance the force the utterer actually needs to do more rather than less.[10] In other words, although not every act of thinking is an act of asserting, the latter is constitutive of and explanatorily prior to the former—This is an important modification of the Frege Point. According to this modified Fregean picture, the force and the content is explanatorily indissociable even though distanceable. An unasserted content is still different from other intrinsically non-intelligible materials, because a distanced thought is, nevertheless, primarily forced (asserted) and thus demarcated as intelligible.

As discussed in Chapter 3, the differentiation of asserted contents from unasserted ones is not a regular sort of classification but rather exemplifies an essentialist model, according to which the asserted contents and the unasserted contents constitute an asymmetrical pair that refers to the singular genus for all species of conceptual contents, of which the asserted content is the primary species. Correspondingly, we can say that the act of asserting is the primary species for the act of thinking in general, and the act of distancing implies qualification and is hence explanatorily secondary.

---

[10] Though I prefer using the notion of 'distancing' in my dissertation, the notion of 'cancellation' is more usually used in the literature. Cf. Hanks (2015, 90ff.) and Recanati (forthcoming).

Now, we can combine this modified Fregean picture with the Austinian picture of illocutionary act: The act of assertion, as said, is equivalent to the illocutionary act; and similarly, the act of distancing is equivalent to the locutionary act. (Shown below.)[11]

	primary (committal)[12]	*secondary* (non-committal)
Modified Fregean Distinction	assertoric	unasserted
Modified Austinian Force	illocutionary	locutionary

Fig. 4.1.

As Fig. 4.1 shows, we have an asymmetrical pair of notions for thinking: *Assertion* (or the illocutionary act) and *distancing* (or the locutionary act). While the assertion stands for the paradigmatic case of thought-presentation, the distancing indicates a non-central or qualified case of thought-presentation. The distancing, in other words, is secondary to and asymmetrically dependent on the paradigmatic case of assertion.

It needs to be immediately pointed out that the dependency of the distancing on the assertion is logical and explanatory, rather than causal or behavioral. If I utter a sentence that presents or performs the content *p* without asserting that *p* (as the simplest example: if I utter ~*p*), I do not have to *first* assert that *p* and *then* distance myself from asserting it in two separate acts[13]—The reason should be obvious if we adopt the essentialist model of force–content relation: If the act of assertion and the act of distancing are two separate acts that must be carried out respectively, we would end up with two different kinds of force. Instead, we should regard them as acts essentially of

[11] Hanks expressed the same idea on comparing the illocution–locution difference to the assertion–distancing (what he calls 'cancellation') difference: "Substitute 'assertoric' for 'illocutionary' and 'act of cancelled predication' for 'locutionary act' and you have my account of the actor." (Hanks 2015, 96) I agree with him that the act of propositional thinking in general is inherently assertoric. Cf. also Strawson's celebrated inaugural lecture: idem (1969).

[12] Discussion of commitment see next section.

[13] Moreover, not only should we not regard the assertion and the distancing as two separate types of acts, neither should they be unified under any higher-order act of thinking. Although such a 'semantic ascent' way of thinking is possible, it would yield obvious regress if we must claim that when I utter of a distanced sentence *p*, I thereby, for example, make a higher-order affirmative assertion about my act of distancing (i.e.: I need to assert the distancing of my assertion that *p*, etc.).

the same kind: making an assertion is performing an illocutionary act without qualification, while presenting a distanced content is performing an illocutionary act with qualification (i.e. a locutionary act). In this respect, assertion and distancing refer to the different constitutions of the act of discourse.

I will discuss in the next section how a correlation between the Fregean and the Austinian picture could be developed into a positive act-theoretic approach to assertions and to discursive acts in general. Before that, in the rest of this section, I would like to clarify why I also call my treatment of the Austinian performative force a 'modified' Austinian view.

I call my treatment of the Austinian force also as 'modified' for two reasons: The first reason has already been explained—namely I adopt a broad reading of assertion that is equated with illocution—In this sense, the illocutionary acts are essentially truth-evaluable. The second reason is that I adopt a slightly modified reading of Austin's locutionary act. As said, Austin listed three kinds of locutionary act: phonetic, phatic, and rhetic. But in speaking of locutionary acts there is an important difference between distanced thoughts presented in rhetic acts on the one hand, and the merely phonetic acts (such as a scream out of pain) or the merely phatic acts (such as the computer outputting a sentence) on the other hand. It signifies that while a rhetic act, by default, takes place at agential or personal level, a merely phonetic or merely phatic act is rather sub-personal or impersonal. When Austin declares that all locutionary acts—phonetic, phatic, or rhetic—are *eo ipso* illocutionary acts, he is claiming that not only should we regard the rhetic acts as illocutionary acts with qualification, we should also regard the phonetic and the phatic acts in this way. While I basically agree with his idea, I think the distinction between the rhetic and the other locutionary acts is important—To regard a merely phonetic or merely phatic act as discursive, we need some special treatment. (I will discuss in §5.5 that a phatic act, like a fictional sentence, can be presented as intelligible in a shifted context.)

Readers might recall that when discussing different models for essential pragmatic condition of assertion, I summarized the Austinian model as follows:

(*A-model*) SENTENCE ⟨*distancing*$_2$⟩ CONTENT ⟨*distancing*$_1$⟩ ASSERTION

Note that according to the speech-act-theoretic approach, a content or a sentence is also inherently an act, too: Just like 'assertion' may denote both an act of asserting and the

content asserted, 'content' may also denote both a locutionary act and the content presented by the locutionary act (call it 'locutionary content'). Furthermore, I argued that for Austin there are not really two steps in this model—Fundamentally, there is only one asymmetrical pair: illocution on the one hand, and locution on the other—Hence, while assertion indicates illocution, both the 'sentence' and the 'content' above are locutions. However, if we further differentiate the locutionary acts, we can nonetheless see that 'content' roughly corresponds to the rhetic act, while the phatic act (and perhaps derivatively: the phonetic act) falls under the category 'sentence' above. Later in §4.2, we will see that such differentiation reflects the distinction between qualified (with 'distancing$_1$') and disqualified (with 'distancing$_2$') acts of discursive commitment.[14] In this manner, we can rewrite the Austinian model as follows:

(*A-model**) PHATIC ACT ⟨*disqualif.*⟩ RHETIC ACT ⟨*qualif.*⟩ ASSERTION

According to my modified approach, this is to emphasize that among the various kinds of locutionary acts, only the rhetic act is intrinsically agential and discursive.

4.2 Constitutive Discursive Agency

4.2.1 The Basic Idea of Constitutive Agency

As indicated earlier, the act-theoretic approach combines the (modified) Austinian idea of the illocutionary force in the speech-act theory with the (modified) Fregean idea of the assertoric force. We can do so because the (modified) Austinian illocutionary force is meaning-constitutive and demarcates the content just as the Fregean assertoric force does. Both indicate the constitutive, pragmatic unity of our discursive acts. Call such unity *discursive agency*.[15]

---

[14] If without specification, in what follows when I speak of 'locutionary acts' I refer to the rhetic, i.e. qualified, acts.

[15] To be committed to a discourse is to be a *discursive agent*; and the unity of the agent is her *discursive agency*. But why do we need to introduce the term of agency or agent? The term 'discursive agent' is better than terms such as 'speaker' or 'utterer' because an illocutionary act does not necessarily involve phonetic or other specific expressive communicative acts. We do not need to hold on to the dogmatic understanding that a speech act requires someone who really 'speaks'. Just as we discussed earlier, even though a robot may be capable of 'speaking'

Since illocutionary acts are also called by Austin as performative utterances, equating assertion with illocution also coincides with my suggestion in Chapter 2 to treat assertions as performatives: Assertion is a special kind of performative act, namely, the act that presents the content as true. With this regard, performatively presenting a content as true can be seen as an act of *discursive commitment*—i.e., a commitment to the content's truth. In short, according to our commitment-based view, *a speaker makes an assertion iff she makes a performative commitment to the truth of the asserted content* (while to say that *p* is true—as argued—means nothing other than saying that *p* is discursively appropriate). Therefore, to say that speech acts are inherently illocutionary or that thoughts are inherently assertoric is also to say that they are inherently committal.

Just like that a non-mathematical content always has a constitutive pragmatic force, it always has a discursive agent. A thought is always *someone*'s thought—Here 'someone' could be either an individual person, a group of persons, or an institution—I will argue in §4.3 that the content asserted by an agent can thus be seen as the agent's *doxastic content*. Furthermore, identifying an agent's doxastic content is *attributing* the corresponding discursive commitment (or equally, doxastic commitment) to that agent.

In making an assertion "… is true", the agent commits herself to the discursive appropriateness of the asserted content in its context; and just like the performative speech of a promise, the agent's assertion is a description of the commitment as well as the very undertaking of that commitment by the same token. Therefore, the commitment has two aspects: From the aspect of the agency, the commitment indicates that the agent

---

phonetically or phatically, it is not really a discursive agent since it never makes any discursive commitment. By contrast, it is fully legitimate to regard a silent language-user—such as *le penseur*—as a discursive agent who 'thinks out loud', because he is capable of making commitments about what he believes in. The crucial thing, therefore, is what the agent is self-knowingly and discursively committed to, rather than his overt expressions as such. (Recall Martin Luther's famous quote: "you are not only responsible for what you say, but also for what you do not say.")

Moreover, we are hence equating speech acts with acts of thinking based on their central, paradigmatic case, i.e., the case of illocution/assertion. Even though not every act of thinking is an act of asserting—and not every speech act is illocutionary—to be a thinker (or an utterer) is inherently to be an agent of discursive commitment, namely, to be one who asserts. From this perspective, the unity of the language-user's act of thinking consists in the language-user's very agency of discursive commitment—just like that the unity of a thought consists in its performative force of assertion.

is *authoritative* in ascribing the content of the commitment to herself in its context, while from the aspect of the content, the commitment indicates that the agent is held *accountable* for the discursive propriety of the content following universal evaluative standards—Although the act-theoretic approach attempts to explain the demarcability of meaning in terms of the agent's performance, the performance, as an act of commitment, is not an arbitrary one (like swearing with an outburst of anger). Rather, just like that the assertoric force makes no contribution to its content but only constitutively signifies the content's unity, the discursive agency equally signifies the constitutive unity of its act as an instantiation of relevant *discursive norms*[16]—Call the norms instantiated by an act of discursive commitment its *constitutive norms*.

According to this view, whether an act is discursively appropriate (i.e. whether it succeeds in instantiating relevant discursive norms) equally reflects whether its discursive agency is properly constituted. This idea can be traced back to the so-called *constitutivism* in moral philosophy, which links the norms that prescribe us acts we should do with the norms that determine our make-up as agents who perform the acts. Though originally proposed in discussions of moral theories, the idea of constitutivism is helpful to our current act-theoretic exploration of the constitutive unity of content, which, I think, must defend the following two main theses:

> (α) A speech act is an assertion iff it is subject to the norms for discursive appropriateness constitutive of the content it performs.
>
> (β) Whether a norm is constitutive of a content is demarcated by the agent's discursive commitment to follow that norm in performing the content.

Thesis (α) establishes the correlation between the evaluability condition of an act of assertion with the evaluability condition of its content. In other words, it says that an act is an assertion iff it is held accountable of the truth of its content. Thesis (β) is equally

---

[16] Since discursive norms are norms for the discursive appropriateness of one's assertion, it might be natural to think of them as norms that are directly linked to logicality and theoretical rationality, e.g. norms for consistency, accuracy, unbiasedness, non-fallaciousness, conscientiousness, etc. Nevertheless, since our approach to speech acts is not restricted to the narrowly understood 'theoretical judgments' but concerns the act of thinking and discourse in general, the discursive norms actually encompass norms in all aspects of our rational life—Especially, if we incorporate practical deliberations into our generally discursive approach, its norms may include all kinds of ethical norms such as the norms for justice, honesty, diligence, etc.

important, which maintains that the pragmatic force of commitment, rather than being tantamount to content, signifies constitutively the unity of the content by demarcating its evaluability condition—that is, by demarcating its constitutive norms.

With this, we can see how the Austinian illocutionary force could correspond to the Fregean conception of the assertoric force in its response to the demarcation question: As a meaning demarcator, the Fregean assertoric force neither adds anything extra to the content, nor functions as a regular classifier, but only presents the content as true. Now, through the utterer's discursive commitment, we can see that the same goes for the Austinian illocutionary force—It neither renders the content private nor introduces any specific mode of speech; rather, it indicates a commitment to presenting the content as nothing but true—As shown, according to Austin, to assert is to perform the content in manner that the agent does nothing more than in the illocutionary line. The agent's discursive commitment in the performative act is therefore *sui generis*, just like the Fregean force: There is no more general kind of commitment that applies to cases other than discourses. (E.g., there is no commitment to talking in sleep.) Commitment is only possible when it presents a conceptual content as conceptual content.

Still, we want our commitment-based approach to also account for the distanced analyzability of conceptual contents—Though all speech acts are inherently committal, there could be qualifications that distance the force from the content and perform the content in a non-committal way. But it remains so far unclear how the constitutivist view could explain non-committal acts—It means that even though one might agree with the act-theoretic view that the committal act and the non-committal act are an asymmetrical pair that refers to the genus of all discursive acts, the following problem persists: According to which standard can we determine in a discourse whether the agent is committal or she is actually distancing or inhibiting her commitment?

Peter Hanks suggests that the 'inhibition' or the 'cancellation' of the force generates a 'cancellation context', and we can identify cancellation and therefore demarcate a non-committal act through such cancellation context.[17] This is apparently true, but it also immediately demands further explanation: It is true because I can only find out whether one 'really' asserts or unasserts that *p* by affixing further contexts to his words—For example, he might later disavow the assertion by saying that "I only

[17] See Hanks (2015, §4.1).

said 'if'", or might reassure me that he meant to assert 'literally' that *p*, or even 'literally literally' that *p*, and so on—However, it should also be obvious that having recourse to context alone (i.e. by invoking further content implied by the context), at best, does not help solve the problem, and at worst, undermines the possibility of eventually determining the meaning of any thought, since there can be no guarantee when a complete context could be obtained.

As I will argue in Chapter 5 in more detail, the theses about constitutive agency must not be question-begging: namely, it is the commitment that demarcates a content, not the other way around—More detailed discussion has to be postponed to later chapters, where I will argue that the distancing of a commitment consists in a *change* of context, by which the content's constitutive norms are also changed—Here, it suffices to point out that just like that the assertoric force cannot be reduced to a predicate, we equally cannot reduce the commitment (or its inhibition) to a specific property of the agent. Rather, the commitment signifies the very *constitution* of the agent, so that a discursive agent is inherently committal. It means: First, the disambiguation of a sentence's status of commitment consists in the condition of attributing commitments to the relevant agent; And second, the attribution of discursive commitment is inherently self-attribution—in the sense that a commitment is an agent's *self-constitution*.

Regarding the first point: if we want to demarcate a non-committal speech, or to determine its 'cancellation context', we need to identify a distanced agency. Consider the example that an actor delivers Prince Hamlet's famous soliloquy on stage: In this case, we attribute the doxastic content and the corresponding status of discursive commitment expressed by that soliloquy to Hamlet but of course not to the actor himself. But when we are aware that Hamlet's lines are non-committal, we are thereby identifying Hamlet as a distanced agent 'embedded' in the actor's speech (qua oratio obliqua). In other words, we are aware that there is a distance between the actor's own doxastic commitment and Hamlet's. Moreover, we see that there is an asymmetry between the two: The actor *himself* is a discursive agent (agent without qualification) first, and an acting Hamlet (agent with qualification) second.[18]

[18] Strictly speaking, performing a line on stage fictively is still different from an embedded indirect speech since a fictive speech requires a subordinate, shifted context. But even in this scenario, we can still regard the actor himself as being committed to a global discourse, namely, as a discursive agent whose illocutionary utterance is in a sense to report Hamlet's doxastic content. In fact, the identification of distance is always secondary and reflective—Globally

But how do we then identify someone as a committed agent in the first place? It leads us to the second point: the attribution of commitment is inherently self-attribution. It means that we can identify the commitment of an agent without qualification only 'from within', i.e., in terms of *first-person thought*.

Let me summarize the idea of the asymmetrical distinction between committal and non-committal acts as follows:

THE CONSTITUTION OF AGENCY

CENTRAL	**Committal Act** (Assertion) *Condition*: No further condition but the absence of qualification *Content*: First-person thought
DERIVATIVE	**Non-committal Act** (Distancing) *Condition*: Various qualifications *Content*: Disagreement

Fig. 4.2

To explain this framework, I shall first illustrate in §4.2.2 and §4.2.3 some basic ideas about the qualifications for performing discursive commitments, and present in §4.2.4 a slightly modified version of the framework. I will say a bit more on doxastic content, first-person thought, and discursive commitment attribution in §4.3. And finally, in Chapter 5 and 6, we will discuss the possibility of disagreement (qualified discursive act) and how it is dependent on first-person thought.

4.2.2 Qualifying and Disqualifying Conditions for Commitment

When we speak of qualifications or qualifying conditions, we always speak with reference to certain imperfections: If *S* is *F* without qualification, it brings us no further knowledge about what a qualifying condition for being *F* could be. Therefore, when we ask what is the condition for the agent to succeed in performing a discursive commitment, we need to know under which conditions the agent would *not* succeed. Indeed, this is also Austin's own approach in considering the conditions for successful speech acts—He started by considering the "things that can be and go wrong", which he

---

speaking, we must first understand that *Hamlet* is a play before we get to understand what is happening on stage. Cf. §6.2.2.

called *infelicities*.[19]

Austin distinguished two kinds of conditions, the failure of satisfying any one of which will induce infelicities in a speech act—If the condition of the first kind is not satisfied, the speech act is said to be a *misfire*: "when the utterance is a misfire, the procedure which we purport to invoke is disallowed or is botched: and our act […] is void or without effect."[20] On the other hand, if the agent failed to meet the second kind of condition, his speech act is said to be *abused*: "[when the utterance is abused,] we speak of our infelicitous act as 'professed' or 'hollow' rather than 'purported' or 'empty', and as not implemented, or not consummated, rather than as void or without effect."[21]

Austin succinctly characterized the difference between these two cases by his admirable mastery of the English language. The central idea, I think, is this: When the utterer misfires, no speech act has really been performed. The utterer might have 'said' something—such as in sleep-talking—but he did this not in a self-knowing manner; or, the utterer might be disposed to perform a speech act, but suffered from confusion and thus failed to perform any, e.g. being baffled. On the other hand, when the utterer abuses, his speech act has been performed, though not appropriately.

Consider Austin's example of the speech act of 'marry'. If the utterer says "I'll marry you tomorrow" while sleep-talking, or in a total bewilderment—imagine the utterer, while shocked (or manipulated), speaks unconsciously "I'll marry you tomorrow(?)"—his act of speech misfires because what he says is neither true nor false, but intrinsically empty of any meaning. However, if the utterer says "I'll marry you tomorrow" intentionally yet nevertheless does not make it happen the second day, his promise is hollow though not empty of meaning. He has performed a speech act, but what he says is false.

I think we can borrow Austin's distinction between 'misfire' and 'abuse' in our discussion of the conditions for discursive commitment: If the utterer's speech act misfired, no discursive commitment has been made. In this case, if we can still call the utterer a discursive agent, then the agent was only pathologically constituted, and his act is only pathologically discursive; correspondingly, we can call such kind of condition

[19] See Austin (1975, 14).

[20] Ibd. p.16.

[21] Ibd.

for the non-success of performing commitments *pathological condition*. Under a pathological condition, the utterer failed to make any commitment and hence failed to act—The utterer is, in a broad sense, a 'pathological agent'; but in a strict sense, he is *disqualified* from being a discursive agent. By contrast, under a condition where the utterer abuses his speech act, his discursive commitment is alienated or distanced, i.e., its performance is defective (or as I will discuss in later chapters: it implies disagreement); but nevertheless, his act is still discursive. In short, while a pathological act is disqualified from being agential, a non-pathological, non-committal act is *qualified* for being agential.

These two kinds of conditions reflect the ambiguity of the notion 'condition' (and 'qualification') in the English language. In one sense, if *S* has a condition for being *F*, *S* failed to be an *F*—Pathological conditions are in this sense; i.e., they are *disqualifying conditions*. In another sense, if *S* has a condition for being *F*, *S* is still an *F*, but less absolutely so—Conditions that implies disagreement are in this sense; i.e. they are *qualifying conditions*.

If we distinguish two kinds of conditions for performing discursive commitments, we can similarly distinguish two kinds of evaluative norms for the pragmatic success of performing commitments: On the one hand, there are norms that are effective in purely pathological cases; and on the other hand, there are norms proper for the discursive appropriateness of the contents performed by the discursive act. For our investigation, the latter kind of norms is more important, because only this latter kind of norms could be constitutive of the discursive commitment of an agent—On the contrary, in purely pathological cases, since the agent failed to achieve any discursive commitment, the norms are not constitutive of his act and therefore not constitutive of his (pathological) agency[22]—They are not discursive norms proper but rather external threat, coercion, or enforcement towards the pathological agent.

To sum up, I distinguish qualifying conditions (henceforth QC) from disqualifying conditions (DQC) and suggest that the conditions for non-pathological, non-committal acts are QCs, not DQCs. Both two kinds of conditions are important and cannot be confused with each other. My approach thus differs from the radical, constitutivist view of agency that maintains that whenever an agent's commitment is

[22] I will give more discussion on constitutive norm in Chapter 5 and 6.

distanced, that agent immediately ceases to exist (other than being pathological)[23]—Such a view rather describes purely formal operations of, say, arithmetic: Under the postulates of ordinary arithmetic, adding 2 to 2 always yields 4. Once an operation results in 5, instead of being a qualified operation of addition, it is, strictly speaking, not an addition at all—As I argued in Chapter 3, in such case the purely formal agency and its content do not exemplify a relation of essential generality. In contrast, in non-purely-formal cases, it must be possible for us to conceive of a non-pathological agent as a qualified agent.

Consider that Jack is orienting himself on a map, which requires his knowledge of navigation. Suppose, then, that he makes a mistake in the orientation.[24] The first thing we want to identity is whether Jack is pathological. If Jack is an infant or is drunk, he would not be accountable for his 'mistake in orientation' since he was not really doing that—Even though he, if drunk, could be globally accountable for having allowed himself to get drunk in the first place—By contrast, only if Jack is *not* pathological, we can say that he has made a commitment, which is qualified (invokes disagreements) for anybody who realizes the mistake: In disagreeing with Jack, we consider his discursive commitment as falling short of satisfying the discursive norms, i.e., the norms for properly exercising the knowledge of navigation.

Yet again, the agent is only accountable for his failure or falsehood when he is primarily identified as a committed discursive agent (I will discuss this in detail in Chapter 5). It is only in this case does a disagreement, or debate, or persuasion, or correction, make sense—Though disagreeing or debating with a drunken man is not impossible, neither is it intrinsically discursive—What is crucial in this picture is that an agent, unless pathological, is never deprived of any of his commitment. By disagreeing with Jack, we thereby distance ourselves from his assertion, or maybe he could distance or inhibit himself from a commitment of his and turning non-committal on it. But being

---

[23] Cf. Robert Nozick's humorous caricature of what he calls the 'coercive philosophy': "perhaps philosophers need argument so powerful they set up reverberations in the brain: if the person refuses to except the conclusion, he *dies*. How's that for a powerful argument?" (Nozick 1981, 4). More discussions in Chapter 5.

[24] Here I use the example of mistake and falsehood to discuss the non-committal act for simplicity reason. Of course, not all non-committal acts imply falsehood. The scope of non-committal acts is more general: it includes any act of distancing or alienation of the agent's commitment.

non-committal does not make a commitment vanish or disappear, it only adds more qualifications. Being non-committal, in short, is being committed with qualification.

4.2.3 Qualified and Unqualified Acts of Commitment

So far, we have spoken of the conditions for commitment only in a negative way—only when there are certain defects. But this way of talking about the conditions for commitment seems to be unhelpful when it comes to identifying an *unqualified* agent and its commitment.

Indeed, we need to be especially careful here: Admittedly, we can assume that there are certain 'conditions' that the agent needs to 'satisfy' to perform a commitment. But in this sense, the 'condition' is not a condition *for* making a commitment, which an agent may or may not satisfy. In other words, it is unhelpful to say that there are 'conditions' for an agent and his committal act of assertion that the agent needs to do more to 'satisfy' in order to unqualify. For a discursive commitment *as such*—i.e. a commitment that is not distanced or alienated—its 'condition' is not a qualification, but rather the *absence* of qualifications. Otherwise it would be like attaching the clause of *ceteris paribus* to every assertion—Imagine Jack saying: "By going east, *ceteris paribus*, I will reach the post office shown on the map." This seems to be equivalent to saying that *p* if *not non-p*, which is not only unhelpful but also makes the whole assertion empty.

Yet still, a discursive commitment is not a divine act. It is either appropriate or inappropriate, and hence it must be evaluable in terms of its appropriateness. In other words, it is not yet completely senseless to say that there must be certain conditions for a commitment—even an unqualified one—to be what it is. An appropriate act of commitment is not an act out of nowhere, but an act *out of reason*. In act-theoretic terms, such reason can be called the agent's *motivating reason*. But is it contradictory to what we just claimed, namely that the commitment is an act without qualification? If the agent's motivating reason is not a condition *for* making the commitment, how could we understand it after all?

At this point, we are seemingly captured in a dilemma similar to the rule-following problem. (It is also similar to what Kant has presented in his *Third Antinomy*, namely an apparent contradiction between the thinker's spontaneity and Free Will on the one hand, and the causal determinism of the laws of nature on the other.) That is, the

notion 'motivating reason' appears to be an oxymoron: on the one hand, since an unqualified agent must be 'self-motivating', it seems that if the agent has a motivating reason then it must be 'unconditional'; yet on the other hand, it seems that for the motivating reason to be a 'reason' at all it must be answerable to further norms and conditions.[25]

To better comprehend the problem, let us consider the QCs for a commitment. The QCs for a commitment are reasons or discursive norms according to which that commitment could be evaluated or analyzed. For example, Jack might be either to the east or to the west of the post office he wants to get to. They are reasons *pro tanto*—weighing options or evaluative standards—if we regard Jack's act of commitment as an act to best explain his location. Comparing to the reasons *pro tanto*, we can say that the motivating reason for an agent's commitment is the reason *pro toto*, namely, the conclusive, all-things-considered reason. But now, we may find that the phrases 'pro toto' and 'all things considered' appear to be uninformative: If we stick to the viewpoint that the phrase 'pro toto' expresses an unrestricted quantification of conditions (pro tanto reasons), we would obviously be unable to perform a discursive commitment after all, since no non-divine being could ever put 'all things' into consideration for making an assertion.[26]

Our reasons for acting are always defeasible: In the example that Jack wants to determine his location. Under ordinary circumstances, he might have two weighing options that he could be either to the east or to the west of the post office. But the list of options and conditions could actually go indefinitely long—He could be standing at the north pole, so that he is neither to the east nor to the west of the post office. Or it could be that his map is out of date and the post office no longer exists any more, and so on.

---

[25] Also, cf. Terry Pinkard's formulation of what he calls the 'Kantian paradox' of autonomy in idem (2002) chap.2.

[26] Accurately speaking, acts of deliberating or weighing over options and drawing conclusions are acts of inference rather than assertion. But since the conclusion of an inference, after all, is an assertion, I think it is not far-fetched to compare between 'asserting out of reason' and 'inferring from reasons'—While 'asserting out of reason' emphasizes the order of explanation, 'inferring from reasons' emphasizes the order of recognition. Further discussions cf. §6.2.4 and Appendix 2.

The agent's conclusive reason, therefore, cannot be assimilated to a reason pro tanto.[27] Pro tanto reasons are non-committal reasons—They are norms the agent may consider to follow or not to follow—By contrast, an agent's reason pro toto is committal.

From this perspective, we can see that the way the motivating reasons are to the pro tanto reasons is the same as the way the asserted contents are to the unasserted contents. What is true about the assertoric force must also be true about the motivating reasons: It means that the motivating reason is always explanatorily prior to the pro tanto reasons of a commitment. The motivating reason is not assimilable to a weighing condition or set of weighing conditions from which the commitment can be inferred (like the *ceteris paribus* clause); rather, it has a unique pragmatic implication related to the agent. Moreover, such a unique pragmatic implication must remain truth-evaluable. Our task, then, is to explain what it could be that 'motivates' such a reason—There should be neither an oxymoron nor anything mysterious in the notion of motivating reason. For instance, it does not indicate anything like the 'unmoved mover' or the initial 'unconditioned condition' for all other conditions.

4.2.4 Discursive Power and Constitutive Norms

The unique pragmatic implication of an agent's commitment that motivates the performance of the commitment, I suggest, is the agent's *discursive power*. The agent's performance, in other words, is the manifestation of her discursive power. To begin with, let us distinguish two kinds of agency without qualification:

a) The Divine Intellect: agent without qualification in an absolute sense, i.e., agent without any qualification whatsoever.
b) The Finite Intellect: agent without *known* qualifications.

Intuitively, we should focus on *us*, namely the finite intellects. As finite intellects, our discursive agency is unqualified only *by default*, and can be cancelled or alienated by qualifications. For us to make a commitment is to constitute ourselves as an agent—with knowledge of our discursive power; yet when we perform a commitment, the absence of qualifications is not absolute but socio-historically conditioned, because our agency signifies the constitutive unity of the discursive norms that are implicit in the

[27] Beside, if we interpret the reason pro toto as a higher-order reason *for* selecting certain pro tanto reason against others, the argument would be either regressive or empty.

our previously known QCs.

In this regard, we can accordingly modify the framework of the constitution of discursive agency as follows: ('→' means 'implies')

THE CONSTITUTION OF (FINITE) AGENCY

CENTRAL	**Committal Act** (Assertion) *Condition*: No further condition but the absence of known QCs *Content*: First-person thought / Self-agreement → Discursive Power
DERIVATIVE	**Non-committal Act** (Distancing) *Condition*: Various known QCs *Content*: Disagreement → Discursive Norms

Fig. 4.3

The content of a committal act of assertion is a first-person thought because it implies that the agent is conscious of her own act of commitment and thus knows her assertion in a unique, first-handed, and authoritative way—It is a same act asserting a content and asserting the appropriateness of so asserting. Hence, we can say that the agent in assertion agrees with herself on the propriety of the content asserted.[28] But the confidence in her self-agreement consists in her discursive power, which is constituted by the norms realized socio-historically in the agent's practice of thinking. According to the view of constitutive agency, these norms constitutive of the agency are precisely the norms constitutive of agent's assertion.

The idea of the internalization of norms shown in the framework above will be explained in later chapters, and discussion on how the internalization of norms could be seen as manifestation of the discursive power has to be postponed until §6.3.[29] At this point I only would like to make some clarifications about my use of the term 'power':

The discursive power is a power of discursive commitment, namely, a power of spontaneous and autonomous rule-following. Equally, it is the power of thinking and knowing in general. However, it would be problematic to say that such a power is a

[28] I will say more on first-person thought in §4.3.

[29] The most important idea, as I will argue, is that the agent's self-agreement must be understood as a dynamic, self-educating process consisting of the internalization of norms.

power *to φ*, e.g. a power *to assert p*.[30] Otherwise, either the manifestation of the power is described as being 'exhausted' by the description of *φ*-ing so that there is no spontaneity in that power (like the 'power' of a stone to fall), or the manifestation of the power is conditioned or qualified by another power so that the power is heteronomous (such as a car's power to move)—I will offer a bit more discussions on this in §6.3—My main idea is that it is crucial for the understanding of an autonomous and spontaneous power that it is neither sheer potentiality nor signifies any specific function.

Hence, when we speak generally of the discursive power as a power manifested in the act of assertion, we should bear in mind that it is not a power *for* assertion, but a power *in* assertion—Similarly, if we allow ourselves to adopt terms like 'the power *of* assertion', or 'the power *of* Reason', etc., it might be useful to distinguish two senses of the preposition *of*: the *of* equivalent to the *genetivus obiectivus* (e.g. the slaughtering of an ox) and the *of* equivalent to the *genetivus subiectivus* (e.g. the mooing of an ox). And we should see 'the power of assertion' as exemplifying a case of genetivus subiectivus.

The discursive power cannot be seen as a power *to φ*. But we still have various specific powers in our lives, e.g. the power to swim, the power to write a philosophical article, the power to juggle four balls, etc. These are powers that are okay to be put in the formula 'the power to *φ*'. Hence, to distinguish these specific powers from the discursive power, in this dissertation I will stick to the term 'capacity' instead of 'power' in denoting these specific powers—Thus, instead of saying 'the power to swim' I will say 'the *capacity* to swim'.[31]

The specific capacities correspond to the specific knowledge and skills we have. And to evaluate these specific knowledge and skills we need various evaluative norms. For me to write a philosophical article I need pertinent philosophical skills, which is governed by the norms for qualified philosophical articles. For Jack to orient himself, in

---

[30] On the one hand, it could be okay to say that we have power to think, to assert, or to know. But 'to think' or 'to assert' is not an act if without content. Hence such saying is rather empty. On the other hand, however, it makes no sense to speak of a 'power' to assert *that* snow is white—Our powers and capacities are always generic. (We learn to discern *a* hornbeam instead of *this* hornbeam. Similarly, it is senseless to say Sophie has the capacity to discern *this* hornbeam.)

[31] Moreover, if we express these specific capacities using the word *of*, such as "the capacity of swimming", obviously it exemplifies a case of genetivus obiectivus.

turn, he needs knowledge for navigation, which is governed by the norms for correct navigation. In other words, these specific capacities are capacities to follow the relevant norms. Moreover, since we are finite intellects, in exercising our capacities there could always be norms that we failed to follow, which introduce qualifications into our knowledge and skills. When we were born, we were basically ignorant of all these norms, and we live by building up our knowledge and skills so that we will not fail to follow those norms in future.

For an infant, the norms for writing philosophical articles are *external* to the constitution of his discursive agency, because he could never be committed to the writing of a philosophical article and therefore never be committed to following those norms. Similarly, I might purport a commitment when I am drunk or delusional—I might say, for example, that I am "committed to writing a philosophical article right now". But those are, clearly, empty words. In fact, I failed to make the commitment—my performance of commitment 'misfired'—because I failed to constitute myself as a discursive agent with the capacity of following the norms for writing philosophical articles. In this case, those norms are not constitutive of my discursive agency, the same way as to the infant. They are DQCs for me to write a philosophical article, rather than QCs.

A central idea for the commitment and norm-following is that the agent not only acts in accordance with the relevant norms, but also acts in a knowledgeable manner, i.e., with the knowledge that his act's accordance with the norms is precisely *because of* his capacity of acting in accordance with them. It could be possible for me—with a very small probability—to type out a brilliant philosophical article while drunk just like a monkey randomly hitting keys on a typewriter. But the article is brilliant not because of my capacity to make it so—In a word, the agency of commitment must be anti-luck. As I will discuss further in Chapter 5, to distinguish a commitment from a coincidence or a guesswork, the relevant norms must be *internal* to the constitution of a discursive agent, in the sense that they are constituted as a motivating reason for the agent to form a belief and assert accordingly.

To conclude, it must be able for an agent to express his motivating reason not only in form of '*I ought*' but also in form of '*I can*'.[32] This is why we need the notion of discursive power—Accurately speaking, for an agent to make an assertion, she must be

[32] Cf. Rödl (2013) and Zhan (manuscript).

committed to following certain specific discursive norms, which implies the exercise of her relevant capacities in a knowledgeable manner. This, in turn, consists in her discursive power, which signifies her act as an act of spontaneous and autonomous norm-following. A goldfish has the capacity to swim, and it may, in one way or another, be subject to the norms for elegant swimming. But it never exercises its capacity to swim in a knowledgeable manner; namely, it never has any know-how to swim. To have the 'knowledge how' it requires the agent's manifestation of her discursive power, which constitutively signifies the agent's knowledge of the exercise of her capacities in a discursive act. Therefore, the discursive power can be seen as that which 'internalizes' the norms in an act of discursive commitment: It is what 'motivates' the norms and constitutes them in a unity of discursive agency. Without such 'motivation', it makes no sense to say that an agent gets to exercise any of his specific capacity or follow any norms—In fact, without such 'motivation' we would be unable to distinguish an agent's capacity from a sheer potentiality or functionality, and one's act would be lifeless, contingent, or heteronomous—This brings us back to the basic idea of constitutive agency introduced earlier, which contends that the norms constitutive of the content are also constitutive of the very agency performing that content. But now, we have a mediating concept in this picture, i.e., the concept of the discursive power—The discursive norms do not immediately constitute an agency. Rather, they acquire their constitutive unity only through the agent's implicit discursive power. The constitutive norms are only those that are 'internalized' by the discursive power through the knowledge of the agent's previous QCs.

Here are two examples about the agent's discursive power and constitutive norms:

Example 1

Imagine that my apartment is on fire, and I shall jump out of the window into the river below. I am committed to doing so as this is a motivating reason. (Admittedly, my actually jumping into the river is not a purely discursive act but a practice with lots of physical actions going on, but the justificatory core of those actions consists in the discursive act in my mind.) I can, furthermore, inhibit my commitment and run through the weighing options in terms of pro tanto reasons—Should I jump into the river or stay in the room, or should I even sit down and start to write a philosophical essay instead?

Etc. All the pertinent norms (e.g. norms that govern the evaluation how dangerous it is for me to swim in the river or stay in the room) are internal to my agency because they are within my capacity. By contrast, it is *not* a pro tanto reason whether I should jump out of the window to fly in the sky, because unlike jumping, swimming, sitting, and essay-writing, I do not have any 'knowledge how' to fly. Therefore, even though there are countless norms out there for flying, say, norms concerning aerodynamics and air traffic safety, those norms are not internal to my agency in this case. This shows how, as a finite intellect, my discursive power—and therefore my agency—is socio-historically conditioned. I am discursively agential when considering reasons for and against sitting down and writing an essay when my apartment is on fire—If I decide to jump into the river, then I regard sitting down and writing an essay as a QC for my discursive agency, of which I speak in a distanced way (in the subjunctive)—By contrast, it would be a DQC for my discursive agency if I consider whether I should jump out and fly in the sky even though it is impossible for me to do so.

Example 2

Here is another less 'practical' example: Suppose that Sophie is asked to sort three trees into two kinds. While the trees look rather similar, Sophie wants to try her best. She finds that tree A is bigger and has greener leaves, while tree B and C are smaller, with less green leaves. Moreover, the trunk of tree A is straighter and apparently stronger than B's and C's. Hence Sophie decides to sort tree A into one kind, and B and C into the other. Suppose, however, that if any botanist was asked the same question they would do the opposite—They would always sort A and B together, since they are American hornbeams, while tree C is a European hornbeam—To the tutored eye, so to speak, the difference of tree C from the other two is so significant that the seeming peculiarity of tree A, which is caused only by different conditions of nutrition and age, becomes trivial and negligible—According to my interpretation, then, both Sophie and the botanists have performed 'appropriate' (i.e. qualification-free) discursive commitment since both of their acts of assertion followed their constitutive discursive norms. The skills and norms of telling American hornbeams from European hornbeams apart is not constitutive of Sophie's agency since she has never been trained for this.

And her assertion is only false, say, when tree A is in fact not bigger and its leaves are not greener.[33]

4.3 Commitment, Belief, and First-Person Thought

4.3.1 Commitment and Belief

So far, we have introduced many aspects in charactering the distinction between the qualified and unqualified acts of discursive commitment (e.g. asserting vs. unasserting, committal vs. non-committal, reasons pro toto vs. reasons pro tanto, etc.). But there is still one important aspect that needs to be discussed, namely the agent's *belief*.

One important feature of a doxastic content is that its parts or arguments may not be *salva veritate* substitutive. An agent who holds a belief about Phosphorus may not hold the same belief about Hesperus. Or, for our example earlier, Sophie holds a belief about the tree in front of her (say, "We have such a tree in our town.") may not hold the same belief about American hornbeam (i.e. "We have an American hornbeam in our town."). These are different thoughts because they are different commitments with different implicit discursive norms.

Yet on the other hand, we should not 'privatize' beliefs—In so far that beliefs are conceptual contents, they must be universally sharable. In holding a belief, I do not thereby turn it into some kind of idiolect. I neither dictate or own its truth, nor, in standard cases, need to think myself as standing in a relation to that belief—Certainly we can report a belief, but it is not like as if we are reporting a property of ours.[34] Beliefs are always thought and communicated in terms of sharable reasons. Just like Matthew Boyle argued, when we question someone's belief—

---

[33] In a non-expert context it is usually of little significance to classify trees or distinguish between jaguars and leopards, or between alligators and crocodiles—In fact, the standards for predicative accuracy vary dramatically from context to context. More discussions to follow in §5.6.

[34] An example of this latter approach is the Russellian Multiple Relation Theory of belief-ascription, according to which the agent is considered to be an essential component of a proposition. Russell proposed this idea in order to solve the problem of the objective meaning of falsehood and individual senses: Generally speaking, '*S* believes that *p*', as a complete proposition, states a relation between an agent, an n-tuple, and an attitude, somehow in the form of F(*S*, *p*, b).

> We do not ask him "How do you know that you believe P?", as we might ask him how he knows that his blood pressure is elevated. Indeed, we do not normally entertain any question about a person's epistemic relation to his own beliefs at all. We normally address our questions and criticisms entirely to the soundness of the propositions he believes, criticizing them or the grounds he gives for them. (Boyle 2009, 125)

I think it points out two things. First, the identification of beliefs must take place at the agential, namely, personal level, because holding a belief requires the agent's self-knowing act of commitment—It implies the agent's expression of self-agreement and cannot take place in a merely subpersonal or impersonal level—And second, the agential control we perform on our beliefs (i.e. the assertion) is not an extra act to be done to the belief-states somehow we possess (e.g. as an installation); quite the contrary, it is *the* act for us to hold or possess a belief. In Boyle's words: "Our beliefs are thus not normally things on which we act; they are themselves our acts." (Boyle 2009, 119)—Having a belief *is* performing agential control on it.

A crucial point in this regard, therefore, is that we should not confuse belief with memory—Precisely speaking, we must reject a 'storage' or 'installation' view of beliefs, namely the view that beliefs are propositions that somehow 'stored' or 'installed' in our mind or brain, and is recalled or reiterated only when we need to. The reason is similar to the reason we reject the view that contents are inherently force-free—We should reject the view that doxastic contents can be entertained as inherently force-free content dissociated from the act of judgment with their truth-condition and unity remaining self-contained:

> It is as well to reserve 'belief' for the notion of a far more sophisticated cognitive state: one that is connected with (and, in my opinion, defined in terms of) the notion of judgment, and so, also, connected with the notion of reasons. (Evans 1982, 124)

If we accept Evans' suggestion—and I think we should—then we have to distinguish the distinctly cognitive act of believing (in the sense of judging and reasoning) on the one hand from the mental act of 'recalling' something previously 'stored' in memory on the other hand, which does not intrinsically involve judging and reasoning. In this sense, a pure act of recalling something would be like reporting one's psychological history. Nishi Shah and David Velleman once argued as follows:

> If the question is *whether I already believe that P*, one can assay the relevant state of mind by posing the question *whether P* and seeing what one is spontaneously inclined

> to answer. […] But the procedure requires one to refrain from any reasoning as to whether P, since that reasoning might alter the state of mind one is trying to assay. Hence asking oneself whether P must be a brute stimulus in this case rather than an invitation to reasoning. (Shah and Velleman 2005, 506f.)[35]

Shah and Velleman were right in claiming that a pure act of recalling something from memory would not involve reasoning but rather be a "brute stimulus". But in such case, we must also no longer say that the subject is reporting a *belief*. The reported content would be a non-linguistic one (e.g. recalling a feeling) or a linguistic one but only in a sense that is not intrinsically meaningful (e.g. recalling the sentence a stranger has said to me in an unknown language the other day). To report *p* as a belief of mine, on the contrary, essentially carries my awareness of my commitment and status of conviction.

Just like that a conceptual content always has a constitutive force, so does a doxastic content. Therefore, rather than reporting or iterating 'stored' propositions, belief is also performative. In fact, as we said that the paradigmatic act of presenting a thought is the act of judgment / assertion / discursive commitment, I think it fully applies to doxastic contents, i.e., we identify doxastic contents precisely in terms of discursive commitments. For me to consider whether I believe that *p* is nothing but to consider reasons for and against the truth of *p*. Therefore, my belief implies my constitutive commitment to the appropriateness of the assertion that *p*;[36] and attributing a doxastic content to an agent is also precisely attributing discursive commitment (or equally, *doxastic commitment*) to that agent.[37]

---

[35] Cf. also Boyle (2011).

[36] Admittedly, in real cases we do often need the function of memory (or what Tyler Burge calls the function of 'content preservation') to help us recall beliefs. To recall or reaffirm a belief, so to speak, we do not always need to reexamine or rehearse through all the evidence and grounds that we used to have for adopting that belief at the beginning. I trust my previous judgment as part of my knowledge and acquisition. But in this sense, the memory is no longer a "brute stimulus" but rather a more sophisticated, virtue-oriented mechanism of *self-trust*. This shows that our memory could be intrinsically cognitive and participate in the socio-historical constitution of our discursive agency, which makes up the default-appropriateness of the doxastic content. Cf. Burge (1993). Cf. also §5.4.2 on the distinction between 'self-trust' and 'self-reliance'.

[37] Since my interpretation of agency is not constrained to individual human beings, the interpretation of belief also applies to collective beliefs, i.e. beliefs of a group of persons or an institution.

It might seem that I can have a belief without asserting it. For example, if you ask me: "Do you believe there are a million trees on Earth?", my apparent answer would be "Yes", despite the fact that I have never explicitly made that assertion or commitment before you ask me that question—But this exactly proves the point: beliefs are not some stored propositions on which we operate. There is no difference whether you ask me "Do you believe *p*?" or "Can you affirm / assert *p*?"[38]

Another approach that tends to dissociate assertion from belief is claiming that one can make an assertion without believing it. Such an approach allows one to purport or pretend an assertion. A typical example is social constructivism, which I will discuss in §5.4. For me, however, although one can of course purport an assertion and hence be doxastically non-committal, it cannot be the central case but is only derivatively intelligible.

To put it in a nutshell, I suggest that according to the act-theoretic, commitment-based view of constitutive agency, there is an immediacy between assertion and belief, which shows us that it would be problematic to distinguish between the act of assertion as an act of *forming* or *altering* a belief on the one hand, and the *knowledge* or the *holding* of that belief on the other— People indeed tend to adopt such a distinction in their act-theoretic approaches to assertion: For example, Grice used to claim that the intention of assertion (the 'M-intention') is an intention to induce a certain belief in the hearer. While it is possible for a view of 'belief-inducing' to accommodate the view of discursive agency as depicting the dynamic processes of reason-giving acts in teaching and disagreeing, the view itself taken at face value is problematic precisely because it suggests the distinction between having and controlling beliefs—The problem of such view is not only that the assertion is not essentially hearer-oriented, but that it risks resulting in a doxastic voluntarism, according to which the agent might be able to intend to induce a belief in himself without yet having that belief. Instead, according to the

[38] Moreover, the inferential consequence of the claim is also part of my belief or assertion. My assertion / belief that there are a million trees on Earth implies the assertion / belief that there are 557238 trees on Earth—as long as the norms of relevant arithmetic knowledge is constitutive of my commitment. There is, however, a difficulty here: if we regard one's beliefs as a set of sentences, this set may not be deductively closed—In believing *p*, I may not, in fact cannot, believe in the infinite set of logical consequences of *p*, such as (*p or q*), (*p or q or r* ...) etc. Hence we should differentiate between a default consequence guided by constitutive norms (as conversational implicature) and a merely derivative consequence of a belief. More discussions on belief sets and belief revision in §6.2.1.

view of constitutive agency, we must maintain an immediacy between belief and assertion, because for a non-inhibited performative act of commitment, it hardly makes sense to say that we need to distinguish an 'active' exercise of agency (e.g. undertaking a commitment, settling a question, making up one's mind, achieving a conviction, etc.) from a 'static' status of the agency (e.g. the status of being committed or convinced, holding a belief, etc.).

4.3.2 Commitment and First-Person Thought

Earlier I claimed that an agent's commitment is identified inherently only 'from within', i.e. in terms of first-person thought. Now, we can see that the first-person thought consists exactly in the immediacy indicated in the previous subsection: I suggested that belief is a performative just like assertion; namely, it is a same act to hold a belief and to assert it. (To unhold a belief, in turn, is to unassert it, i.e., to distance oneself from the commitment, which is derivative.)[39] If I assert *p*, in declaring the truth of the content that *p*, I also immediately declare the truth that I hold *p* as my belief (that I know that *p* is my belief). Since believing *p* is nothing but attributing the commitment to the truth of *p* to myself, i.e., asserting that *p*, it follows that it is the same act to believe *p* and to know that I believe *p* (or to assert *p* and to know that I assert *p*).

This 'performative immediacy' enables us to explain how it is intrinsic to the nature of commitment that the agent makes the commitment self-knowingly. We can call it the 'Transparency Condition' of the self-knowing agent. The Transparency Condition contends that that *I think p* primarily adds nothing to my assertion *p*. In other words, the assertion that *I think p* does not primarily report a preexisting statement in my psychological history, as if the content *p* is an indirect, reported speech; neither does 'I think' introduce any specific attitude other than the performative force. The Transparency Condition thus rejects treating the agent and its propositional attitude—like the Russellian Multiple Relation Theory suggests—as independent parameters of a proposition. The 'I think', just as the performative force, cannot be assimilated to a

[39] We have similar immediate connection in the problem of rule-following: It is a same act to adopt or recognize a rule and to follow it. (To *not* follow a rule—e.g. to treat it merely as a reason pro tanto—is actually derivative.).

predicate.[40]

If we regard the first-person thought here as yielding a kind of self-knowledge, then this kind of self-knowledge is transparent and hence infallible—But the self-knowledge here is transparent and infallible only in the sense of performative immediacy—In fact, this is how Anscombe famously characterized the infallibility in the intention in acting:

> If a person says 'I am going to bed at midnight' the contradiction of this is not: 'You won't, for you never keep such resolutions' but 'You won't, for I am going to stop you'. (Anscombe 1957, §31)

Of course, the content of one's assertion is fallible—it may be discursively appropriate or inappropriate. But Anscombe's argument is about the practical immediacy in the structure of intentionality, just like the performative immediacy in assertion and belief. Since the performative force is not a predicate—Anscombe claims that intention indicates the agent's non-observational knowledge—it cannot be falsified by any opposite predicate, and there is hence no space for falsehood between my assertion *p* and my transparent self-knowledge expressed in that assertion, namely that *I think p*.[41] Moreover, since the force is not a predicate, there is equally no space for any 'positive' predication between one's assertion and her transparent self-knowledge thereby expressed in the assertion. In our context, the self-knowledge hence does not pick out or refer to any determinate 'self'—For this reason, Anscombe has yet another well-known saying that 'I' is not a referential term. I think her claim is justified at least to the extent that the self-knowledge is not knowledge applied on a specific kind of object.

Such view is coherent with our picture of the constitutive doxastic agency: The believer knows of her beliefs not by any further inner observation but by *being* the believer. Since the assertoric force indicates the constitutive unity of the content, the

---

[40] Irad Kimhi has been doing some significant work recently on the related issue (See Kimhi forthcoming). He claims that there is no categorematic difference between *p* and *I think p*. (Rather, they make a so-called 'syncategorematic difference'.) I cannot discuss his work here, but at least I agree with him on the negative point—To use Kant's words, the 'I think' makes no empirical contribution but indicates the synthetic unity of consciousness.

[41] I think this kind of performative immediacy also exemplifies and defends what Evans called the immunity to error through misidentification; namely, it is constitutive of the discursive agent who makes an assertion that she must be immune to error through misidentifying the content of her assertion.

assertion or the doxastic content already *is* the agent's transparent self-knowledge—In Kantian terms, every thought is explicitly or implicitly uttered by the underlying 'I think', while the unity of the thought is equally the unity of the agent's consciousness—There is thus no room for the agent to add further conditions (such as the ceteris paribus condition) or attitudes to her assertion. For example, although there can well be cases when Sophie is uncertain about the credibility or the truth of a statement—e.g., she might rather prefer a probabilistic estimation of the degree of certainty of it—it, however, only means that in such cases the statement subject to uncertainty does not make up the whole content of Sophie's assertion and commitment. In this sense, we may say that the doxastic content always takes the 'widest scope' (since it corresponds to the unity) of the agent's consciousness.[42]

If we understand the transparent self-knowledge, we can also understand why the act of attributing commitments or doxastic contents must be inherently *self-attribution*: On the one hand, if being committed can be seen as an act of self-attribution, we must *not* regard the self-attribution as a derivative kind of the act of attribution as such—I do not first identify a primitive commitment with its target of attribution still pending, and then somehow get to decide to apply the supposedly primitive attribution-apparatus on myself—There is no such primitive apparatus and there are no two steps. And on the other hand, to be sure, 'distanced attributions' of commitments must also be possible, namely, it must be possible to attribute a belief or commitment to someone other than myself. But we need to see that even in such a case it must also already imply *that one*'s act of self-attribution. It is in this sense that Castañeda introduces his quasi-indicator (*), to show that my report about Sophie that *she* thinks p* shares a same structure of first-person thought as '*I think q*', so that although the content *p* is an indirect speech and non-committal for me, it nevertheless is a direct speech and committal (and therefore a doxastic content) for Sophie. Attributing a doxastic content or commitment to an agent, therefore, must be inherently a presentation of the content as the agent's direct speech. For example, we can speak of Sophie that *she* thinks p*, while *p* = (Tree B should be sorted together with Tree C.); and we can also say that *she thinks q*, while *q* = (This American hornbeam should be sorted together with the European hornbeam.). But in the latter case, the quasi-indicator

[42] Uncertainty should be differentiated from confusion—In the case of confusion, the agent is rather pathological and unable to constitute a unity of content or consciousness on its own.

can no longer be used—We are doing something more than attributing a commitment to Sophie, i.e. we are attributing a commitment to Sophie in a qualified manner (i.e. with distancing or alienation).

4.3.3 Performative Immediacy and Thought *De Se*

In the previous subsection, I argued that for me to assert that *I think p* it primarily says just the same thing as the assertion *p*. (Or similarly: that Sophie asserts that she* thinks that *p* primarily says just the same thing as her assertion *p*.) I suggested that the immediacy in this structure of first-person thought shows its infallibility and hence a form of transparent self-knowledge. But there is a right way and a wrong way to understand the infallibility and the transparency of self-knowledge: The right way is the Anscombean one introduced earlier. The wrong way of understanding, by contrast, is the Cartesian anti-skeptic approach to self-knowledge, which is based on the self-certainty of the subject's introspective observation (i.e. the so-called 'internal theater')—So far, we have no resource to prove this latter understanding.

The immediacy we have argued for is a performative immediacy of the self-knowing (or self-constituting) agency, which shows that as long as the 'I think' indicates an act to perform the content *p* as the agent's doxastic content, it must be constitutive of the agent's very act of asserting that *p*. The 'I think', therefore, says nothing other than the sameness between the agent's assertion and her belief. In Kant's words, the agent understood in this way rather expresses a 'transcendental subject', which gets us no knowledge about a determinate 'self':

> Through this I, or He, or It (the thing), which thinks, nothing further is represented than a transcendental subject of thoughts = *x*, which is recognized only through the thoughts that are its predicates, and about the which, in abstraction, we can never have even the least concept […] (A346/B404)

For Kant, the problem of the Cartesian picture is that it takes the existence of the thinking being as a 'real predicate', while it would be a transcendental illusion to take any knowledge to be at the same time infallible and real.

However, such an Anscombean-Kantian treatment of self-knowledge seems to be too 'thin'—The self-knowledge in this sense brings about no concrete knowledge *about* the agent. And even though we know that it is constitutive of an agent's assertion *p* that *she* thinks p*, it does not help us single out any particular person who makes this

assertion. But we are not just transcendental subjects; rather, we are persons or individuals that are, in one sense or another, empirically determinable: On one hand, we need to use the first-person pronouns as distinguishable from pronouns of other persons; and on the other, even though they occur in different assertions (such as the two 'I's in "I am YZ." and "I have leg pain."), they must be unifiable into an individual 'self'.

It might seem, therefore, that though the performative immediacy between assertion and belief is able to explain the transparency between the agent's self-knowledge and the content of her assertion, we somehow still need a further, different account of immediate self-knowledge that is able to yield some determinate knowledge or thought *de se*, i.e., that which predicates something *of* ourselves.

In this respect, Richard Moran distinguishes in his book two kinds of immediacy, which he calls 'practical' and 'epistemic' immediacy, respectively.[43] The 'practical immediacy' is similar to the performative immediacy between assertion and belief. The 'epistemic immediacy', in turn, refers to the immediacy in what he calls the "introspective awareness", which enables us to make reiterated reports on our beliefs e.g. *I think I think p*, etc., which nevertheless enjoys transparency and infallibility.

But how can we defend the transparency and infallibility of such an introspective awareness? One possible explanation is interpreting introspection in the Cartesian sense; but as argued, we are unable to defend any determinate self-knowledge in this way—Moran himself also rejected the Cartesian picture at the beginning of his book—If we force to let the introspective sentence to express some determinate *de se* thought, e.g. saying that I think it is *I* (namely YZ) that think *p*, etc., then I have thereby asserted more things, namely, I am not only the asserter of *p*, but also YZ—I thereby take myself to be a real predicate—But then there is obviously space for fallibility and misidentification (e.g. Oedipus' mistake).

Another possible explanation is that the transparency is based on the immediacy between the agent and her committal status. Supported by the 'epistemic immediacy' as such, however, the reiterated 'I' in the assertion remains thin and still does not introduce any determinate knowledge *de se*. In other words, such an explanation collapses the distinction between the 'epistemic immediacy' and the 'practical immediacy', since, as argued, it would be problematic to distinguish between the agent's 'active' exercise of

[43] See Moran (2001, 131ff.).

achieving a conviction on the one hand and her supposedly 'static' status of being convinced on the other.

The (Anscombean) Kantian view correctly denies that the performative 'I think' could guarantee us privileged access to any determinate knowledge *de se*.[44] However, the Kantian view indeed has serious limits and needs to be expanded or revised. If we trivialize the knowledge *de se* and insist that self-knowledge must always be read as a thin constitutive expression of the transcendental subject, then we would end up with a deflationism, according to which the performative 'I think' is superfluous. I will give a more detailed discussion on the limits of the thin reading of performative agency in the next chapter, where I discuss some potential objections to the constitutive agency.

[44] John Perry, for example, proposes such a view of the privileged access of the first person, which, he contends, is expressed as ineliminable essential indexicals. Cf. my discussion in §5.3.3.

CHAPTER 5

Agency and Constitutive Norms

5.1 A Worry About Constitutive Agency

According to the act-theoretic approach sketched in the last chapter, the demarcation of content is inherently demarcation of the norms constitutive of the content, which consists in the speaker's constitutive agency in performing the commitment to the content's truth or discursive appropriateness by instantiating those norms—In short, the act-theoretic approach explains the interpretation of a content as being constituted by the speaker's agency or the pragmatic, performative force. However, if we cannot say anything peculiar about the agency or the pragmatic force, it seems that such an explanation might be simply tantamount to the semantics of content.

In her essay "Two Kinds of Agency", Pamela Hieronymi argues that there could be a "radical interpretation" of agency, according to which "the 'normative status' cannot appear apart from an exercise of agency"[1]—It is quite obvious that the constitutivist approach outlined above also defends such a view—Nevertheless, though admitting her sympathy with the view, Hieronymi claims:

> This more radical interpretation will obviously require positing an exercise of agency in a surprisingly wide range of cases, and so, one might think, either will be wildly implausible or else will require an objectionably deflationary account of agency—agency will be attributed wherever we find an attitude with a certain sort of 'normative status,' regardless of whether we find, there or in the agent's history, any discernible mental processes or activities that we could independently identify as an exercise of agency which we might associate with that attitude.[2]

---

[1] See Hieronymi (2009, 142, fn. 4).

[2] Ibd.

While I do not think the use of the notion of 'agency' is any more 'abundant' than notions of, say, 'attitude', neither do I think the approach to constitutive agency must be simply 'deflationary'[3]. Proposing a view of constitutive discursive agency is to account for the assertoric force as indicating the constitutive unity of content. With this regard, I proposed in §4.2.1 that we must be able to defend the following two theses:

> (α) A speech act is an assertion iff it is subject to the norms for discursive appropriateness constitutive of the content it performs; And—
> (β) Whether a norm is constitutive of a content is demarcated by the agent's discursive commitment to follow that norm in performing that content.

Thesis (α) corresponds to Frege's Insight into the assertoric force that it must be meaning-constitutive and it does not introduce any additional feature or property to the content. The reason for an agent to assert a content must be immediately answerable to the norms for discursive appropriateness of the content performed by the agent's speech act. It thus resists treating the force to be semantically inflationary. However, Thesis (α) taken alone seems to be just a mutual definition between content and act. And the worry is that if without further explanation, a biconditional statement as such is insufficient since it does not yet tell us whether any side of the biconditional really holds—One might say that it says nothing but two hypothetical statements, neither of which can have the validity of its premise independently understood. An account of constitutive agency based solely upon Thesis (α), so to speak, would be indeed captured by an 'objectionable deflationism', i.e., the mutual definition between the norms intrinsic to the asserted content on the one hand, and the norms intrinsic to the performance of illocution and the constitution of the agent on the other.

Instead of adhering to the abstract notion of agency as such, we need to see that the agency of a finite intellect implies the agent's discursive power, which is always mediated by socio-historical qualifications. The notion of discursive power is to open up the possibility for a *pluralism* of constitutive norms of discursive agency. It seems, then, that the burden of explaining this pluralism and a possible escape from the 'objectionable deflationism' falls on Thesis (β)—How do we know whether a norm *is* a valid and authoritative norm for evaluating the appropriateness of an assertion? Answer:

---

[3] In what follows I use 'deflationary' or 'deflationism' in the sense of the "objectionably deflationary account" that Hieronymi refers to.

Only when the norm is constitutive, and whether a norm is constitutive is demarcated by the agent's discursive commitment—An agent commits herself to follow the norms that are within her power, because it is her discursive power that signifies the norms' constitutive unity. (For example, the botanical criteria for taxonomy are external to the evaluation of Sophie's assertion about sorting trees, because they are not signified by Sophie's discursive power as a norm constitutive of her assertion.)

Nevertheless, it is not yet very clear what Thesis (β) and the notion of discursive power can really do—As I have already indicated in the last chapter, there is a basic worry that it remains unclear, above all, how the recognition of failure and qualification could be possible if everyone is just making their own assertions basing on their own constitutive norms—If we cannot give a satisfactory account of this, we will not be able to understand what a constitutive norm really is; and if we do not know what a constitutive norm is, the whole theses of constitutive agency will become rather senseless.

In sum, we see that for an act-theoretic approach the following two questions are interrelated:

> (*A*) What is the constitutive norm of a type of act (in our case: the act of assertion)? And—
> (*B*) What is it to (be committed to) follow or to instantiate a constitutive norm in acting (in our case: in performing an assertion)?

Yet most of the times when we try to explicate one of those, the other becomes seemingly inscrutable. Dummett made a revealing analogy with the game of chess to show the problem:

> Formal description of chess could be given by describing the initial position of the pieces, and giving rules for what constituted a legitimate move from any given position. Players make moves alternately, and the game ends when there is no legitimate move. The end-positions are classified into three categories: White checkmates Black, Black checkmates White, and stalemate. […] This formal description suffices for the mathematical theory of chess, and for the formulation of chess problems. In terms of it one can state and prove such theorems as that it is impossible to force made with two knights, or describe a position and ask whether it is possible to arrive at from the initial position by seriously legitimate moves. It will not, however, suffice by itself to provide us with a 'theory' of chess as an activity: it is not enough to tell anyone what it is to play chess. […] From the formal description it is impossible to tell what, in playing

> chess, the players trying to do, namely to produce an end-position which falls into a particular one of the three categories [… without the] information about what constitutes winning.
>
> […] Now, just as it is possible to describe to someone what it is to play chess without presupposing that he understands what winning is, and therefore that he already understands some similar activity so it should be possible to describe the activity of using language without presupposing that it is already known what significance it has to call one class of sentences to class of 'true' sentences and the other the class of 'false' sentences. (Dummett 1981, 296f.)

A formal description of chess defines the legitimate moves in a game of chess. From a formal description, we can formulate a set of rules for the game, and we can call these rules the constitutive norms for playing chess: If you played a legitimate move of chess, you obeyed its constitutive norms; and if you played an illegitimate one, you thereby violated a constitutive norm. However, this is different from being *committed to follow* the norms. A computer, for example, does not need such a commitment.[4] A properly designed computer chess program will be always playing the legitimate moves. But making legitimate moves is not enough for characterizing the nature of the activity of playing a game.[5] (As Dummett indicated, we can well imagine a computer knowing all the 'winning' positions, making all the legitimate moves, but nevertheless avoid reaching one of those 'winning' positions. To us the computer would be playing another game instead of chess. But for computers, it is just a question of classification to distinguish between the set of winning positions and the set of losing ones. It is just a distinction between X and Y. While one can label one 'win' and the other 'lose', or the other way around, for the computers the *concept* of 'winning' does not really matter.) For us humans, we think that one cannot realize the concept of a game without having the 'aim' or the 'intention' of the game. Unlike a computer, I may have made an illegitimate move and thereby failed to follow the rules of chess—but this failure is only intelligible when I am indeed intended or committed to follow the rules in the first place.

---

[4] Dummett himself used the example of Martians.

[5] Perhaps we can say that the totality of legitimate positions corresponds to the extension of the game, but not its intension.

Chess-playing is a type of act. And one way to define a constitutive norm—and thereby answers question (*A*)—is claiming that a norm is constitutive of a type of act iff every performance of the act is subject to the norm. According to this definition, we can thus see the formal description of legitimate moves as the constitutive norms of chess. But as our discussion just showed, even though a constitutive norm defined as such provides a universal standard for *evaluating* the propriety of a performance *being* subject to the norm, it nevertheless fails to provide any standard for *demarcating* a performance as a token act of this very type. Namely, it cannot answer the question (*B*) what is it *to subject* oneself to the norm as a norm constitutive of one's act.

The same argument applies to assertion. The act-theoretic approach treats assertion as a type of act and generally accepts that there are norms constitutive of the practice of asserting. Moreover, advocates of the constitutive norms of assertion claim that the constitutive norm[6] must do two jobs: It should demarcate assertions from other types of acts, and it should provide standards for the assertions' propriety.[7] In the next two sections, I will examine some typical approaches to the constitutive norm of assertion, to show that all of them have the fundamental difficulty to properly address the notion of agency and are therefore liable to objections. In §5.2 I shall introduce the recently standard approaches to the constitutive norm of assertion. I will conclude that they try to explain question (*A*) at the cost of failing to answer question (*B*), and thus are unable to reply to the deflationists' objections. In §5.3 I will discuss a Kantian approach that attempts to answer question (*B*), but I show that it is unable to reply to skeptic challenges.

To respond to these objections, we must reconceive the relation between agency and its pluralistic constitutive norms. But the difficulty is that we need to motivate the pluralism without sacrificing the idea of the autonomy of the agency. After discussing what I call the social constructivism in §5.4, I will focus in §5.5 on the problem of understanding how disagreements and mistakes work. I conclude in §5.6 that we need a 'dynamic' view of constitutive agency to better comprehend how the discursive power

[6] Most advocates of constitutive norms of assertion believe that there should be a singular norm (thus *the* norm) constitutive of assertion. Thus I will adopt this singular definitive use in the discussion that follows. But I think this is improper—Though I support the idea of constitutive norms, I hold a pluralistic view about norms.

[7] Cf. Jennifer Lackey (2007).

'internalizes' norms through disagreements and mistakes. (I will offer a positive sketch of the dynamic view in Chapter 6.)

5.2 The Constitutive Norm of Assertion

In contemporary debates about the constitutive norm (or equally: the constitutive rule) of assertion, various versions of the constitutive norm have been proposed.[8] They include but are not restricted to:

(Truth Rule) One must: assert *p* only if *p* is true.

(Warrant Rule) One must: assert *p* only if one has warrant to assert *p*.

All these versions can be traced back to Williamson's proposal of a Knowledge Rule in his 2000 book:

(Knowledge Rule) One must: assert *p* only if one knows that *p*.[9]

Note that the expression 'One Must' in the formulations above is considered to indicate a wide-scope reading, i.e., the norm should govern the whole conditional. Therefore, we can generalize the various formulations of the constitutive norm of assertion in the following form:

(1) *S* ought to (assert *p* (only if *p* is so-and-so)).

If we adopt (1) as a valid formula for the constitutive norm of assertion, it is then up to debate which standard can correctly fit into the 'so-and-so' description of *p* in (1) as the constitutive norm of assertion. I cannot engage in these debates in detail here—Taking Williamson's Knowledge Rule for example: on a charitable reading, I am indeed sympathetic with his view that knowledge and assertion are intimately associated or perhaps even mutually constitutive (since speech acts are, after all, intrinsically cognitive acts); I also agree that knowledge plays a crucial role for evaluating assertions and comprehending thoughts. But an assessment of the Knowledge Rule requires an investigation into his theory of epistemology, which obviously cannot be carried out in

[8] Cf. MacFarlane (2011). Cf. also K. DeRose (2002), J. Kvanvig (2011), J. Lackey (2007), and D. Whiting, (2015). For criticism of the constitutive norm view cf. e.g. C. R. Johnson (forthcoming).

[9] Williamson (2000) chap. 11.

this dissertation. Similarly, the Truth Rule seems also plausible—given that we can also offer a proper reading of the notion of truth. I thus want to remain neutral on which of these versions of constitutive norm is more appropriate and whether some of them might overlap.

Moreover, it has also been the subject of contentious debate whether the constitutive rule should represent a necessary or/and a sufficient condition in its formulation. Taking the Knowledge Rule for example, some argue that representing knowledge as only a necessary condition for assertion seems to be too weak, while some argue that representing knowledge as a sufficient condition for assertion seems to be too strong.[10] Since the debate on the conditional relations also depends on which version of constitutive norm one choose to defend, I will remain neutral on this, too.[11]

But we still can do something here. Before those debates to take place, it is rather of primary importance to clarify what a constitutive normative condition or constitutive rule for a type of act should mean—as what I shall argue in this section. Our goal, namely, is to assess the general idea of the constitutive norm of assertion as such.

A natural way of conceiving of the constitutive norm of assertions is to say that if *a* is a constitutive norm for assertion, then any act, as long as it is an act of assertion at all, is subject to this norm *a*. In accounting for *a*, one thereby tries to answer the question:

(*A*) What is the constitutive norm for evaluating an (or: any) assertion?

As indicated in the last section, the challenge to such an account is that it is unclear how it can help demarcate whether an act *is* indeed an act of assertion and therefore is subject to the norm. In other words, the difficulty is *not* whether the evaluative standard offered to answer question (*A*) is a fitting one—regardless which version of the standard it would be—but whether it can answer the question:

[10] Debates on whether the knowledge norm specifies a necessary and/or sufficient condition cf. e.g. Brown (2010), Simion (2016).

[11] We need a more careful treatment of the notion that a norm is 'constitutive'—As I argued in §3.3.1, the relation of constitutivity actually says something more than the relations of necessary and/or sufficient conditions.

> (*B*) What is it to perform an assertion and thereby to follow or to instantiate the constitutive norm of assertion?

If the constitutive norm of a type of act cannot demarcate a performance as a token act of this very type, then the constitutive norm can equally not instantiate itself. For example, we might formulate the constitutive rule of chess in a similar vein:

> (2) One must: make a move χ only if the piece(s) for move χ is in such-and-such position.

According to our discussion in the last section, formulations like (2) are formal descriptions of the evaluative standards of (the moves of) chess. Since an act according to such formal description could be just classificatory—like the computers do—such an act could only extensionally or behavioristically coincide with the game of chess, but not intensionally so.

But this is exactly what many have taken the standard reading of the constitutive norm of assertion to be:

> To say that the rule is constitutive of assertion is to say that nothing that is not subject to this rule can count as assertion. (McFarlane 2014, 101)

> In the imperative, assert *p* only if *p* has C. As used here, 'must' expresses the kind of obligation characteristic of constitutive rules. […] The rule unconditionally forbids this combination: one asserts *p* when *p* lacks C. (Williamson 2000, 241)

To differentiate us from computers, it is a crucial point to acknowledge that for an assertion, the violation of its constitutive rule is possible. Otherwise, as Williamson admits, "it would be pointless to forbid it." (ibd.) Therefore, it must be possible for one to be subject to the constitutive rule but nevertheless violate it. It thus seems to suggest that we should better distinguish between the norm applied or instantiated in *evaluating* an act and the norm applied or instantiated in *performing* an act, and maintain that the constitutive norm of a type of act only concerns the former kind of norm but not the latter:

> Constitutive rules do not lay down necessary conditions for performing the constituted act. When one breaks a rule of a game, one does not thereby cease to be playing that game. When one breaks a rule of a language, one does not thereby cease to be speaking that language. (Williamson 2000, 240)

The idea seems to be that if an act of assertion is evaluated as having violated the constitutive norm of assertion, there is still some basic qualified act left for the original defective performance. It is tempting to consider the 'necessary conditions for performing the constituted act' that Williams speaks of to be the locutionary acts (phonetic, phatic, etc.)—If performing an act of assertion is primarily performing a locutionary act, it would seem indeed reasonable to make the concession above that the constitutive norm of assertion does not lay down the necessary condition for performing an act of assertion—Admittedly, accounting for the necessary or the sufficient condition for performing an act appears to be a task not restricted to the scope of conceptual analysis but open to various contexts in experience. For example, when I make assertions I need to be awake, linguistically competent, and perhaps not lose my tongue, etc.—But these conditions are not really norms constitutive of my performance of assertion. In fact, if the constitutive norm of assertion is not instantiated in the act performing the assertion, there would be no way to demarcate an event or an occurrence as an act of assertion in the first place; and as I discussed in the last chapter, the demarcation of an act of assertion (illocution) is logically prior to the demarcation of the accompanying locutionary or other physical acts that materialize the assertion.

If a constitutive rule is instantiated only in evaluating an act but not in performing the act, we may call that rule only *evaluation-normative* but not *use-normative*. There are many cases we can speak of constitutive norms in this way. For example, consider a farmer who got his long-waited spring rain to grow his valuable crops. He might judge happily that the rain came timely. In so saying, he was demarcating and evaluating the event of rainfall in a single stroke—but the norm does not inherently concern the falling of the rain, since there is no consciousness involved in climate to 'gauge' the timely fall for the benefit of the farmer's crops. Though we can still say that there is a norm for the timeliness of spring rain constitutive of the region's seasonal change, the event of the rainfall was not originally governed by the norm—In fact, the event was not even instantiated as an act subject to the constitutive norm until the farmer's evaluation had taken place—I think the concept of natural goodness generally concerns talking of constitutive norms in this only evaluative-normative way.

However, the constitutive norm of assertion cannot be conceived in this way. The discussion in the last chapter shows that assertion as an Austinian act of illocution is intrinsically a self-conscious, autonomous act. It means that the demarcation of an assertion must not be conditioned by any other external standards but the performance

of the assertion itself. But if the constitutive norm of assertion is not instantiated in performing an assertion, there would be no way to demarcate an assertion for future evaluation. Therefore, we can say that in the case of assertion, *the constitutive rule must be both evaluation-normative and use-normative*.

As indicated earlier, the advocates of constitutive norms do maintain that the constitutive norm for a type of act must not only provide a universal standard for evaluating the propriety of a *given* token act but also provide a standard for demarcating a performance *as* a token act of this very type. However, we see that if the constitutive rule of assertion is only evaluation-normative but not use-normative, it is unclear how a performance of assertion could be demarcated in the first place—The formulated constitutive rule of assertion says something like: "If I make an assertion, I am unconditionally subject to the rule which says that 'If I make an assertion *p* then *p* has C.'"—But quite obviously, since the formulation itself does not provide any guidance on how to eliminate the conditional phrases, there is no way of explaining the instantiation of an assertion in terms of instantiating the constitutive rule of assertion which is merely evaluation-normative.

One might argue that I am just trying to knock down a straw man since my example of evaluative norm of natural goodness cannot be what is meant by the advocates of constitutive norms of assertion. Admittedly, there is a difference between natural events and deliberative acts—While rainfall is not an act an agent can choose to do, a game is something one can choose to play or not to play. But my point is exactly that it is unclear how the constitutive rule of an act can demarcate the agent's choice to perform such an act. If this cannot be explained by the constitutive rule of assertion, then there is not really a difference between the norm constitutive of the nature and the norm constitutive of a game, because both the goodness of the nature and the virtue of a game would instantiate themselves based only on arbitrary or external grounds.

To make it clearer, let us consider the chess example again: A computer can 'explain' the logical chain of events: *If* one moves the white queen from d3 to h7, it will create a 'winning' position for white—But such an explanation is neutral on whether the 'winning' position should be pursued. In other words, it does not explain why one "*must*: move the queen to h7 only if it will create a winning position for white". In fact, such a normative requirement of "One Must" is exactly what a formal description lacks—It really raises a question instead of giving an answer: "But why?" "Why should

I create a winning position for white then?" The problem, to use Brandom's words, is that we actually cannot "bake a normative cake with nonnormative ingredients".[12]

A possible way out is to inject some normative ingredient into the description of the winning position, so that the property of a winning position implies the condition of favoring winning over losing. E.g.:

(3) One must: make a move χ only if χ helps create a winning position and one wants to win.

Such a formulation postpones the explanation of the comprehensibility of norm-following to the condition of desirability. In fact, we can even reduce the conditional clause in (3) to a to-infinitive clause:

(4) One must: make a move χ to win.

It expresses a mean-end relation, which can be equally put as:

(4*) One must: win by making a move χ.

Expressions (4) and (4*) represent conclusions of instrumental reasoning. But we can see that such a rule, at best, gives a description that *concerns* or *syntactically binds* the applicability of a norm without being able to explain its own semantic or pragmatic condition of application. Since the desirability of winning cannot be explained by being only presupposed by the rule, a formulation of the virtue of the game of chess (to make such-and-such moves) as a constitutive rule—to use Kant's words—is only *hypothetically imperative*.

Playing a game is exercising one's capacity within a local, instrumental domain. A capacity, I understand, is intrinsically a functional kind. An important feature of a functional kind is that it has evaluative standards constitutive of the concept of its kind: being a ball-juggler means being subject to the norms of appropriateness of the exercise of the capacity of juggling balls; being a hammer means being subject to the norms of appropriateness of the exercise of the capacity of hammering; and being a moth means being subject to the norms of appropriateness of the exercise of the capacities that can

[12] Brandom (1994), p.41, cf. also pp.64ff.

be summarized in the life-form of moths, such as the capacity of metamorphosis, the capacity of flying toward the light, etc.[13]

However, since the instantiations of these functional kinds are externally conditioned, the norms constitutive of the exercises of the capacities are also instantiated by external conditions—Though we can say that a hammer, as a functional kind, hammers, and it is a good hammer iff it hammers well, a characterization as such still falls short of demarcating an instance of hammer from other objects. Similarly, I am subject to norms of ball-juggling *only if* I am instantiated as a ball-juggler in the first place—yet this latter determination is not bound by the norms of ball-juggling.[14] It could be determined by some external motives, such as the desire to look cool, or could be bound by no norm at all, but by deception, coercion, or threat. (Imagine the horrible scene that I am asked to juggle four balls smoothly at gun point or otherwise I will be shot dead. This is what Kant describes as the immorality of treating a person only as a means.) Likewise, to play chess I need to plan and to move the pieces legally, strategically, and wisely, etc.—These are the norms and virtues constitutive of the game. But they are hypothetically imperative since their applicability is conditioned by additional, external motivations to play (and to win) the game.

As I mentioned in §4.2.4, as a manifestation of the discursive power, an assertion cannot be assimilated to the exercise of a regular capacity. Hence a proper view of constitutive agency cannot treat our discursive power as a regular capacity

---

[13] Strictly speaking, these three examples show different kinds of capacities. Ball-juggling is an art or skill, the hammering of a hammer is instrumental, while the capacities of moths indicates the life-form of an animal. But we can see that they basically have the same logical structure in terms of the functional normativity.

[14] The problem not only exists in the cases like deciding whether to play as a ball-juggler, but also in the cases of determining the applicability of a specific life-form to a supposed living-being: For instance, in encountering a moth, I may explain its behavior of flying toward the light by invoking the typical life-form of moths, to understand that it is not contingent for it to act that way. In so doing, I represent my sensation of a flying object as sufficiently grounded, i.e. as exemplifying a universal feature according to the constitutive norms of the life-form of moths. However, this is only based on the condition that what I see *is* a moth so that it is indeed subject to this specific life-form of moths but not others—Maybe it is a non-phototactic insect or a tiny drone that just happens to appear to be moving towards the light—Since it is an empirical question what the thing I see actually is, the representation of a specific life-form can thus only be settled empirically and contingently. In fact, I think this is a major problem of the accounts of natural goodness (typically, Philippa Foot's), namely, the represented goodness is always merely evaluation-normative, but not use-normative.

somehow 'encoded' in us, namely, as a general function like breathing and sweating whose performance is rather instantiated and evaluated by external conditions. It is impossible to demarcate an assertion based on any external evaluative attitude or motivation-attribution without invoking the agent's own constitutive commitment to play the game of assertion and discourse. Therefore, if assertion can be seen as a game at all, it is rather a global, unbounded game, because the agent does not need any extra motivation to play it.

In this section, I examined the typical formulation of the constitutive norm of assertion (Williamson, his approach to knowledge aside). I conclude that such formulation is incapable of answering question (*B*): "What is it to perform an assertion and thereby to follow or to instantiate the constitutive norm of assertion?" But if (*B*) cannot be accounted for, a minimalist or deflationary treatment of discursive agency will seem invincible.[15] To highlight the problem, I have drawn a comparison between the idea of the constitutive norm and Kant's idea of the hypothetical imperative. Thus, it should be unsurprising that the Kantian constitutivists are well aware of this problem. For Kantians, there is a difference between hypothetically formulated norm constitutive of a regular, functional capacity on the one hand, and a categorically formulated norm constitutive of the power of rational thinking in general on the other hand.[16] To apply this idea, the Kantian constitutivists suggest that we can answer question (*B*) by looking into Kant's formulation of the categorical imperative and agential autonomy. For this reason, I want to dedicate some space to briefly examining the Kantian constitutivist approach—that I will do in the following section.

5.3 Kantian Constitutivism

I will proceed in this section as follows: I begin in §5.3.1 by sketching the basic argument of Kantian constitutivism (or what I call *K-constitutivism*). I will introduce a prominent skeptical challenge to constitutivism in metaethics. I argue that the skeptical challenge is somewhat wrongly targeted and has failed to pinpoint the real difficulty of

[15] For a typical, well-developed deflationist view cf. Cappelen (2011), cf. also Cappelen and Dever (2013).

[16] Similar discussions can be found in the argument on the primacy of '*energeia*' over '*dunamis'* in Aristotle's philosophy. I will say more on the manifestation of the discursive power in §6.3.

K-constitutivism. I will discuss the difficulty in §5.3.2, which, I claim, is not restricted to metaethics but concerns the constitution of rational agency in general and its evaluability. I will summarize the problem in our context of pragmatics in §5.3.3.

5.3.1 A Metaethical Debate

For Kantians, there are generally two types of challenges to the rational agency: On the one hand, there is the minimalist or deflationary repudiation of the rational agency, which criticizes that the notion of agency is nothing but a useless definition; and on the other hand, there is the overall skepticism about agency, which claims that there is nothing in thinking and acting that we know of as determinate or objective. Neither of their ideas are desirable—While for the deflationists meanings instantiate themselves rather automatically, for the skeptics we use our language and act fundamentally arbitrarily.[17]

To fend off both challenges, the K-constitutivists adopt an *internalist argument* to account for question (*B*) and therefore our rational agency. Their basic argument is that the reasons constitutive of the content must be understood as the constitutive motivating reasons for the agent's act. The internalist argument, therefore, contends that an agent could not only act merely *in accordance to* the norms and rules, but also act by spontaneously following those norms and rules. And only in the latter case is a norm also a reason, while acting on the consideration of such a norm is acting out of reason.[18]

There are two things to note here: First, most discussions about motivating reasons have been concentrating on specifically practical deliberations and actions. But I believe the topic also applies to the discussion of our act-theoretic approach to assertions in general.[19] Second, the idea that an argument for motivating reason is

---

[17] For discussion on the idea of rational agency in relation to the two problematic, extreme variants cf. Douglas Lavin (2011).

[18] The agency hence must be anti-luck. Such idea can be traced back to Kant, who famously claimed that a person with good will should not only act in accordance to obligation (*pflichtgemäß*) but also 'out of' obligation (*aus Pflicht*).

[19] As discussed in §4.1.2, the account of discursive agency is more general than the specifically practical agency since any knowledge, regardless whether it is theoretical or practical knowledge, as far as it is possible at all, will have to need an account of agency for beliefs and assertions. In fact, many constitutivists, for example Korsgaard (especially in eadem (2009)), and also Kant himself in certain sense, speak of the constitution of a person as an agent of the

'internalist' needs to be carefully handled—Both some prominent Humeans (e.g. Bernard Williams) and some prominent Kantians (e.g. Christine Korsgaard) have taken internalist approaches to the subject's motivating force. But according to the Humean approach, the internal reason is motivating because it expresses the subject's sensible disposition. Such an approach is not a viable option for us, because it precisely rests the application condition of norms and reasons upon arbitrary grounds—To use the Fregean anti-psychologistic argument, my sensible disposition to express a feeling of pain is thinkable only if I am capable of asserting that I feel pain, not vice versa. A content is asserted as a reason not simply because it is something motivating. The so-called internal motivating reasons must be nevertheless sharable thinkable contents, otherwise they would lose their cognitive significance and would not be 'reasons' at all—Of course, I do not expect you to be able to physically feel my pain; but to express a pain as a thinkable content means exactly that if you were in my circumstance you would agree with my assertion about pain, and that the circumstance constitutes a reason for me to assert that I feel pain just as it constitutes a reason for you to assert the same. Therefore, in saying that the constitutivist approach adopts an internalist argument, the approach should be regarded rather as a generally Kantian one, instead of its Humean counterpart.[20]

With these been clarified, let us turn to the typical, prominent skeptical objection to constitutivism. David Enoch's paper "Agency Shmagency" sets a good example of it. Though his paper was originally focused on the constitutivism about practical normativity, the same argument was considered, as he admitted in that paper, to be applicable also to cases of doxastic contents and normative reasons for assertions. Enoch agrees that the central idea of the constitutivist approach is an internalist argument, which conceives the constitutive principle of agency to be able to give the agent sufficient ground for acting accordingly. But he suspects that such argument is in fact ineffective. He assumes that a skeptic could reject the constitutive principle by claiming the following:

> […] Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can't act without aiming at self-constitution, but

---

so-called Practical Reason indeed in this more general sense that covers both the capacity for theoretical knowledge and the capacity for practical knowledge.

[20] Cf. also Velleman (2000).

> why should I act? If your reasoning works, this just shows that I don't care about agency and action. I am perfectly happy being a shmagent—a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of shmagency) of self-constitution. I am perfectly happy performing shmactions—nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of self-constitution. (Enoch 2006, 179)

Since according to Enoch, a skeptic may not care about the constitutive propriety of her act, he draws the conclusion that the principle for the constitution of discursive agency is nevertheless incapable of giving us the motivating force to act.

Enoch is correct to the extent that there is no such over-arching 'principle for good actions' that can be adopted as a meta-level norm for any norm-following act. However, his argument does not form an objection to constitutivism, because according to constitutivism being an agent is precisely *not* an aim. Enoch argues that there must be a further force to get us committed to being an agent, but the constitutivists actually do not need any further motivating force or commitment to make the internalist argument work.[21]

We can see that the skeptical objection to constitutivism implicitly relies on an instrumentalist reading of agency—It contends that the power of performing assertions can be compared to a functional kind, so that the realization of its constitutive norms described in the constitutivist construal of agency—and hence the identification of a discursively powerful agent—always asks for further grounds. According to our discussion in the last section, this is exactly the idea of the hypothetical imperative. If we insist that the hypothetical imperative is the only way to comprehend an act's constitutive norm, it seems indeed hopeless to answer both question (*A*) and (*B*) definitively, i.e. it seems either regressive or circular to maintain that the performance of some sort of act subject to the norm constitutive of that sort is itself an instantiation of that very norm.[22]

---

[21] Discussions on the internalist argument cf. Parfit (1997), Brandom (1998), and Korsgaard (1997).

[22] Hume is perhaps the most prominent representative of this view, who contends that there is an unbridgeable gap between the descriptive and the prescriptive, so that any attempt to account for the intrinsic normativity of our acts would be begging the question. E.g. the following passage (cited through Ford 2011, 100):

However, we have also seen that this consequence is *exactly* what the K-constitutivists attempt to avoid. They precisely reject the instrumentalist reading of agency, and maintains that one does not need—as the skeptics suggest—extra motivating conditions to make the internalist argument work. For the constitutivists, asserting *p* and asserting the discursive appropriateness of *p* are not two separate acts. To be a discursive agent simply *is* to commit myself to the norms and rules constitutive of the discursive appropriateness of the content, i.e., to adopt the normative reasons as my motivating reasons. There is no logical space for us to choose or aim to be discursively inappropriate or to be a shmagent, because making a choice or setting an aim *is* adopting a belief, which is a committal act exclusively empowered by the discursive agency. If I failed to be an agent, I would not be able to make up my mind whatsoever; and I would not be able to think or utter any content after all. Therefore, the K-constitutivist treatment of agency actually excludes the possibility for me to choose to act otherwise or to exercise any other power instead of the discursive power, because the idea of choosing to act in one way or another in fact *presupposes* the subject's constitutive comprehension of her act as an act of discursive commitment.

It is worth noting that up to this point the K-constitutivism agrees completely with my own view introduced in Chapter 4. According to my view of discursive agency, performing an assertion is essentially an autonomous, self-conscious act. It means that asserting a content is equally asserting the discursive appropriateness of the content at the context of use. If so, it should not be a problem to instantiate a norm while performing an assertion, because a constitutive norm of assertion must be constitutive not only of the content but also of the performance of an assertion.

Because K-constitutivism rejects the instrumentalist reading of agency, i.e., rejects treating the discursive power as a functional kind, I conclude that the common skeptical challenge to the constitutive agency (and above all to K-constitutivism), which relies on such an instrumentalist reading, is not effective, since it missed the whole point of K-constitutivism. Admittedly, the K-constitutivism has its difficulties, and I

"To suppose, that the mere regard to the virtue of the action, may be the first motive, which produc'd the action, and render'd it virtuous, is to reason in a circle. Before we can have such a regard, the action must be really virtuous; and this virtue must be deriv'd from some virtuous motive: And consequently the virtuous motive must be different from the regard to the virtue of the action." (Selby-Bigge and Nidditch 1978, 478).

will examine these difficulties in the subsections to follow. But instead of a one-off dismissal of the idea of agency, we should rather examine whether the K-constitutivists have really made a plausible internalist argument to defend an autonomous agency—For this reason, I think the common skeptical challenge is not accurately targeted.

5.3.2 A Deeper Challenge to Kantian Constitutivism

As we have seen in the last subsection, the common skeptical challenge to K-constitutivism draws from the internalist argument a Humean conclusion, which contends that there is no norm general enough to be thought as being constitutive of a type of self-motivating act. However, for Kantian constitutivists, the very idea of a self-motivating act must be governed by a general normative principle—Although the being of a thinker or an agent itself does not express any specific content and therefore cannot be represented as a particular aim or norm, it is nonetheless constitutive of any act of assertion by virtue of being subject to a sui generis principle about the committal assertoric act of norm-following as such—For Kant, such a principle is not *hypothetically imperative* but *categorically imperative*.

According to K-constitutivism, in acting or asserting appropriately, you also determine your own identity as a discursively and doxastically committed agent—In Kant's words, you determine your identity as a person with Practical Reason or the Pure Will—If you do not assert appropriately, it is not only your assertion that is inappropriate; it is also you (the constitution of your agency) that is inappropriate—Moreover, the K-constitutivism is not just proposing a mutual definition between agency and content. Rather, it is the former that indicates the constitutive unity to the latter, not the other way around. A content is demarcated inherently by the agent's self-knowing act of commitment, which Kant characterized as the agent's autonomy (through the categorical imperative). With the help of the idea of autonomy, K-constitutivism is in a position to steer a course to an internalist argument that an agent's motivating reason gives herself sufficient reason to act appropriately—without making our agency either trivial and reducible to content or being subject to an external will.

However, to achieve a feasible internalist argument we still have challenges to face: As mentioned in §4.3.3, the 'I' in Kant's 'I think' is understood as a transcendental subject; namely, it does not denote an object, certainly not an object of experience. Similarly, it seems that the verb 'think' does not express any act of the mind

that invokes differences in space and time. The expression 'I think', in other words, seems to be nothing more than a transcendental 'prefix' to the content. But such a notion of 'transcendental prefix' could be just as confusing as Frege's use of the assertion-sign in his *Begriffsschrift*—Above all, it seems unclear how an anti-deflationary, non-redundant reading of the notion of the transcendental thinker could relate to an individual person, without sacrificing the universal intelligibility of its thoughts.

The K-constitutivists attempt to reject instrumentalism by emphasizing that the instrumental principle cannot stand alone. But neither can the principle of the categorical imperative. A person is neither a divine intellect nor an automaton—If my doxastic content is infinite, there will be no need for identification whatsoever, and my identity will be logically self-evident. But since my doxastic content is finite, my identity is not a logically self-evident identity. Similarly, I do not automatically commit myself to follow whatever norm or adopt whatever reason that anyhow comes across my mind and immediately turn them into the motivating reason for me to perform a corresponding assertion. Rather, I am bound by my particular spatial-temporal conditions. These are not just contingent matters. They take a form that is intrinsic to my all-out judgments and assertions. Towards the end of her book, Korsgaard acknowledged this difficulty but then quickly dismissed it:

> A personal project so characterized clearly does have a private component: you want to stand in a special relationship to something that is good. [… When] you come to see that your contingent practical identities are normative for you only insofar as they are endorsable from the point of view of your human identity, you also come to have a new attitude towards your contingent practical identities. You come to see them as various realizations of *human possibility* and *human value*, and to see your own life that way: as one possible embodiment of the human. (Korsgaard 2009, 211f.)

But what is the point of view of my human identity? How do I come to see some of the contingencies as embodiment of my human identity and therefore as normative to me? At this point, Kantians usually tend to give rather abstract descriptions. E.g. Stephen Engstrom once depicts the legislating role of Kant's Practical Reason in that it "proceeds from the universal to the particular and stands like a kind of Midas to its object: from the highest end to the lowest means, it makes everything it touches like

itself—good if good, bad if bad."[23] This seems to be suggesting that each of us has a hierarchical set of practical identities nesting into one another like a Russian Doll, the highest of which is governed by the categorical imperative. Be that as it may, how does the nesting really work? Namely, how does the constitution of an autonomous thinker relate to a particular subject of commitment who makes assertions and utterances according to her specific competence and educational history implicit in the contexts of those assertions and utterances?

The worry, so to speak, is that it is unclear how the idea of autonomy may account for individual senses. For example: When Christopher Columbus reached the Americas, he believed that he reached the East Indies rather than an unknown New World. Now, we can argue that he was wrong, and that he mistakenly took the land he discovered as the Indies. But as I have argued in Chapter 4, an agent is only accountable for his failure or falsehood when he is primarily identified as a committed discursive agent. Thus, we can disagree with Columbus only if we can primarily recognize what Columbus asserts to be sensical—namely, as being made by an act of autonomous commitment to its discursive appropriateness, rather than being made when he was, say, completely drunk or delusional. In disagreeing with Columbus, we are qualifying his agency rather than disqualifying it; and this depends on our primary understanding that his original assertion that he had reached the Indies was made with the knowledge that he followed the norms constitutive of his agency.

However, the challenge here is that the K-constitutivism does not seem to have enough resource to properly account for the intelligibility of a qualified agency and its individual senses, without which we would be unable to make the claim that Columbus was wrong in thinking that the Americas is the Indies, or that John made a mistake when he said that Hesperus is the brightest object in the sky after the Moon before sunrise. For Kant, what identifies an agent is purely transcendental and thus beyond experience. It is then unclear how disagreements among different agents on a thought could be possible.

Let us consider the following statement:

(5) Columbus thinks the Americas is the Indies.

[23] Engstrom (2009, 91).

The 'Americas' and the 'Indies' represent different senses.[24] In my treatment of discursive commitment, we can say that (5) is a sentence mixed with self-attribution and distanced attribution of commitment: The 'Americas' corresponds to the committal, direct speech that can be attributed to the 'I think' constitutive of the sentence. The 'Indies', in turn, corresponds to the non-committal, indirect speech that can be attributed to Columbus. From this perspective, we might separate (5) into the following sentences in a Quinean style of analysis:

(6) I think *x* is the Americas.

(7) He thinks *y* is the Indies.

According to K-constitutivism, both 'I' and 'he' indicate transcendental thinkers, i.e., whoever that thinks. And analyzing in this way, we translate the indirect speech in (5) into two separate direct speeches. However, with (6) and (7) we are obviously not able to say that they disagree with each other, since each of them has one argument slot left open. If we want to say that (6) and (7) are disagreeing with each other, we must be able to account for the equation: '$x = y$'. But how is this possible? A typical Quinean meaning-holism renders such equation as only contingent (through cross-reference), according to which it would seem trivial to speak of agreements and disagreements among different assertions, since they are always automatically attributed to (and only attributed to) distinct agents (be it 'I', or 'he', or any subject of locutionary act). If so, the account of agency as transcendental thinkers would be no different from a deflationary one after all.

If we could only understand self-knowledge in terms of a deflationary agency, it would mean that whenever we make another assertion, change our mind, revise our existing beliefs, we thereby cease to exist but create another self. And identifying such a subject with Columbus would be also impossible—A proper account of agency must allow a multiplicity of selves, but the knowledge of these selves must *not* be

[24] Or we may say—following Kripke—that (5) concerns both the 'semantic reference' (the Americas) and the 'speaker's reference' (the Indies). See Kripke (1977). In making the distinction, Kripke claimed that when someone says: "The man in the corner drinking champagne is happy tonight", the speaker is only pragmatically but not semantically competent if the man in the corner is actually drinking sparkling water. But for a standard illocutionary act there is no sense to say that the speaker is semantically incompetent (or semantically non-committal). And as I will argue later, we can blame an agent for having misused a term only if he was qualified for being semantically competent in the first place.

'indifferently different'—Instead of the Quinean analysis above, therefore, the constitutivism needs to account for a unifying framework for agreement and disagreement, and we rather need an expression roughly like the following:

(8) I think Columbus, who thinks he* reached the Indies, reached the Americas.

Sentence (8) provides a solution to the equation '$x = y$', i.e. the common referent of my assertion and Columbus' assertion is the continent he* has reached. The quasi-indicator is used here to imply Columbus' self-knowledge—The 'he*' is more than just an indexical; it rather singles out another agent of assertion. And only starting from this common ground could my explanation of Columbus' mistake be inherently sensical—Note that, however, this Columbus' self-knowledge appears to be the content, rather than the subject, of my knowledge. To make it clearer, we can divide (8) into the following two statements:

(9) I think Columbus reached the Americas.

(10) Columbus thinks he* reached the Indies.

To account for my disagreement with Columbus, we need a statement like (10). But according to the treatment of the transcendental thinker, it seems that the pronoun 'he*' is in an awkward position, since it must indicate both a transcendental thinker (because it implies self-knowledge) and a non-transcendental subject (because to maintain disagreement, 'he*' must be understood as an agent accessible in experience, which cannot be assimilated to 'I').

There is, hence, a tension here: On the one hand, the transcendental thinker indicates a subject or agent that lies beyond experience; yet on the other hand, we need an account of agency that lies within experience. We have already mentioned this problem in our discussion on determinate *de se* thought in §4.3.3: It would be awkward, so to speak, to say that we have two kinds of 'I's—namely, a transcendental one and an empirical one—expressed in the following statement:

(11) I think I am YZ.

For K-constitutivism, the two 'I's cannot be confused with each other: The first 'I' indicates the abstract, immediate practical identity or constitution of a self-knowing, unqualified agency as such, which demarcates the sentence as an assertion; by contrast, the second 'I' expresses a socio-historically constrained subject in empirical context. According to such a transcendental picture, in other words, it seems that we indeed have

to distinguish between the two kinds of 'I's: the 'I' of universal legislation on the one hand and the 'I' of token-reflexive expression on the other.

To conclude, since we are finite intellects that err constantly, besides explaining truths in terms of the autonomous agency, the constitutivism must be capable of also explaining individual senses as deriving from the qualifications of our agency at the same stroke. But what a K-constitutivist explanation could be? If we insist, like the K-constitutivists do, that there must be a transcendental, context-independent 'I' constitutive of any performance of assertion, it seems that the 'I think' would be the singular direct speech in the assertion. The rest part of the assertion—the actual conceptual content—would be left indirect and non-committal, insofar as it remains unclear how the autonomous transcendental subject (the 'Midas') get to touch the content and turn it also committal. In short, there is a noticeable gap in the transcendental picture above: On the one hand, the 'I' of universal legislation is always identified from within—but as purely practical agency, which is always unqualified—so that we do not understand its qualifications; And on the other hand, the 'I' of token-reflexive expression is always identified from outside, i.e. through empirical observation, so that we do not understand its intrinsic relation to the purely practical agency of autonomy.

To be sure, Kantians usually do not restrict themselves to the interpretation of agency as a purely transcendental thinker. I would like to leave it here open whether Kant himself or the Kantians has or have succeeded in proposing a convincing account of agency other than K-constitutivism.[25] Nevertheless, I think it is not completely unjust to label the constitutivism caught in tension a Kantian one, for two main reasons: First, constitutivism has its roots in Kant's philosophy, and above all, in his idea of autonomy; And second, it is central to Kant's critical philosophy to differentiate the practical from the theoretical use of the Reason, while the Reason has its constitutive use only when it is purely practical, which has only subjective but not objective validity.[26] If this recapitulation of Kant is correct, it seems that we can understand autonomy also only in

[25] On the one hand, a more thorough investigation into Kant's account of agency requires discussion on his moral philosophy and teleology; And on the other hand, even only focusing on Kant's theoretical philosophy one has to deal with an indisputably rich context, which not only concerns semantics but also epistemology, philosophy of mind and perception. Cf. e.g. Béatrice Longuenesse (1998), and Patricia Kitcher (2011).

[26] Cf. Kant's third *Critique*, §88. Cf. also Zhan (forthcoming).

a purely practical and subjective sense, which, then, gives rise to the distinction between two kinds of constitution of agency: agency qua purely practically autonomous commitment on one hand; and agency qua empirically 'efficacious' act on the other. For this reason, though the K-constitutivists suggest a view of interdependency between autonomy and efficacy, they nevertheless treat the norms of autonomy and the norms of efficacy as essentially different kinds of norms.[27]

5.3.3 The Problem in Our Context

We can see that the K-constitutivism is captured in a tension similar to the Frege's Puzzle, which, as we discussed in Chapter 2, initiated the persistent challenge in contemporary semantics to incorporate first-person and demonstrative thoughts into a compositional semantic framework like the *Begriffsschrift*.

If the Kantian transcendental prefix is not trivial, it seems that we need to force to combine the Kantian transcendental subject of autonomy (i.e. the subject of commitment self-attribution) with the subject of empirical identification, which, as Lewis concluded, means that we will have to accept a rather radical view that all our beliefs and thus all our assertions are irreducibly *de se*, and the transcendental prefix will be understood as a kind of essential indexical.[28] The positive account that Lewis offered is based on the claim that the ascription of knowledge and belief may not be propositional. It suggests that there is irreducible component of the utterer's (indexical) attitude or intention in an assertion / belief. But such an account is obviously solipsistic. It makes the comparison of beliefs across time and across persons impossible.

In Chapter 2, I also illustrated the Kaplanian theory of demonstratives as a prominent proposal for solving the problem. From the perspective of the Kaplanian framework, we can say that both the Kantian 'I' of universal legislation and the 'I' of token-reflexive expression play a context-fixing role.[29] By fixing the context, the *x* and *y* in (6) and (7) will be semantically saturated—In other words, the sentences (6) and (7)

---

[27] For example, for Korsgaard, the norms of autonomy are commanded by the categorical imperative, while the norms of efficacy are commanded by the hypothetical imperative. Cf. Korsgaard (2009), chap.5.

[28] Cf. Lewis (1979) esp. 521f. More discussions see next section.

[29] In pragmatic terms, they indicate the speaker's intention, which Kaplan later called the 'directing intention'. Kaplan (1989b), cf. Perry (2009).

are complete, truth-conditional contents. Their truths or extensions might vary from circumstance to circumstance, but their contents are determinate because both *x* and *y* are understood as 'true demonstratives' that behave like a rigid designator expressing '*that*' continent both 'I' and Columbus refer to.[30] However, if we take the token-reflexive expressions as playing a context-fixing role like rigid designators, the truth-condition of our empirical sentences would appear no longer context-sensitive—After all, since it is an empirical fact that the Americas is the Indies, the mutual-referentiality of *x* and *y* is contingent. This contingency is exactly what we need to discourse on. But according to Kaplan, our discourse and assertions only take place after the contingency is resolved by rigidity. If so, our discourse on whether the Americas is the Indies would be trivial.[31] It would leave us with little choice but to adopt the deflationist treatment of discursive agency.[32]

The persistent puzzle, namely, is how agency could be non-trivial but constitutive of content without turning the content into a private language. As I argued in Chapter 2, we must account for a form of general context-sensitivity freed from indexicality—as a top-down pragmatic condition of assertion rather than a bottom-up one.[33] What is 'token-reflexive' of in a discourse, in other words, should not be seen as consisting in the sheer 'material' side of our language, i.e., as that which needs to be rigidly fixed in a demonstration in order to be forged into the content of universal linguistic intelligibility. Rather, in accounting for general context-sensitivity, we should avoid presupposing a distinction between what is purely material and what is purely linguistic, a distinction usually presupposed in accounts of indexical-use.

---

[30] Kaplan also uses the term '*dthat*' for true demonstratives. See idem (1978).

[31] Cf. also Stalnaker: "In Kaplan's semantics, the paradigm examples of sentences that express necessary a posteriori truths such as 'Hesperus is Phosphorus' are not context-dependent, and so have constant character. [So it] would not help to explain how such statements can convey contingent information." Idem (2004, 298). Cf. also Lewis (1980, 82).

[32] A deflationist rejection of the utterer's intention and self-locating attitude cf. Cappelen and Dever (2013). The deflationist approach, however, avoids the problem of agency at a too high price. E.g., they need to implement a sweeping razor that only allows one kind of necessity in discourse: physical necessity. (As they admit, cf. ibd. p.40.) To me, this is hardly satisfactory.

[33] Cf. MacFarlane's summary of this idea in idem (2009, 231): "As a result of this conflation [of context-sensitivity and indexicality], proponents of contextualism have taken arguments that establish only context sensitivity to establish indexicality, while opponents of contextualism have taken arguments against indexicality to be arguments against context sensitivity".

The Kantian constitutivist approach, as we have seen, is not capable of answering this question. For K-constitutivism, it remains unclear how, as a finite intellect, I could discursively relate my individual being to the being of the autonomous agent described by the constitutivism—Although I do not intend to specifically reject Kant's idea of the categorical imperative, it seems to me nonetheless unclear how the categorical imperative could function as a singular principle for assertion, i.e. a principle of universal legislation that brings certain norms into being effective.[34]

Instead, I think we should view the norms constitutive of assertions as socio-historically conditioned and thus pluralistic, and the demarcation of these pluralistic norms belongs to the implicit side, i.e. the context, of our discourse. In the next section, I shall examine the context-sensitivity at play here. Moreover, we want to understand how commitment can demarcate non-committal and pathological acts, which I will discuss in §5.5. I will suggest a dynamic approach to constitutive agency in §5.6.

5.4 The Social Constructivist Approach to Assertion

5.4.1 Contextualism and its Problem

The evaluability conditions of contents are pluralistic and context-sensitive, which, I claim, are demarcated by the agents' discursive commitments. Hence, we need to clarify what a context is and how an agent's commitment can work context-sensitively. In this subsection, I begin by examining a contextualist approach to the question, to show both its advantages and disadvantages. In the next subsection, I will discuss its rival theory, which I call the social constructivist approach, and argue that it is equally not satisfying.

There are various possible answers to the question: What is a context? For Kaplan, for example, a context is an abstract object. It can be semantically modeled by parameters such as time, place, world, epistemic standard concerning the agent, etc. Stalnaker, by contrast, resists treating context as an abstract object, but contends instead that a context is a set of propositions that are taken for granted as the common ground in

[34] As Kantians admit, the categorical imperative is not conceived as a specific norm or aim, or 'maxim', but a sui generis principle or law. Maybe it is sound to formulate such a principle or law from certain aspect of practical philosophy, but for our problem of assertion-making, having recourse to such a principle as an overarching rule of self-legislation or commitment to autonomy does not seem to me as helpful.

a conversation.[35] I will remain neutral on these approaches to context here—What is important for my purpose, however, is to maintain a generally pragmatic approach to context: Roughly speaking, if we call what is explicitly said by the agent the 'locutionary content', then the context of that content is the implicit information background in terms of which the locutionary content is performed by the agent as illocutionary.[36] A context, in other words, is the essential part of an illocution that represents the implicit condition under which the locutionary content is performed.[37]

With this being clarified, we might want to formulate the context-sensitivity of the evaluability condition of an assertion as follows:

> (*C*) An illocutionary act (i.e. assertion) φ is discursively appropriate iff its locutionary content $\ulcorner \varphi \urcorner$ is discursively appropriate at context *c* in which $\ulcorner \varphi \urcorner$ is performed as illocution.

Or, a shortened version of (*C*) (by adopting the definition of truth as discursive appropriateness):

> (*C**) An assertion φ is true iff $\ulcorner \varphi \urcorner$ is true at context *c*.

But at this point, we need to notice that from the perspective of an illocutionary act, the context of a content is unique and fixed. Since, as argued in Chapter 4, an agent's discursive commitment is nothing but an illocutionary act, we can now have a basic

---

[35] I will discuss the Stalnakerian conception of common ground in §5.4.3.

[36] As mentioned in §4.1, there are at least two possibilities to understand an Austinian locutionary act: as a rhetic act or as a phatic act. Hence, one might argue that the notion 'context' may also receive either a wide or a narrow reading depending on how we understand the explicit locutionary content. For example, if we take a locutionary content to be rhetic, its context might be understood in a narrow sense—the consideration of such a context might be restricted to the so-called 'far-side' pragmatics that determines what is meant beyond saying—By contrast, if we take a locutionary content to be phatic, its context will be understood in a wide sense—the consideration of such a context will also cover the 'near-side' pragmatics such as saturation and use of indexicals. However, since what we are concerned about is the condition of truth and evaluability in general, according to our discussion of the essential, top-down pragmatic condition in Chapter 2 the wide and the narrow readings of context do not form a fundamental distinction here.

[37] Perhaps Lewis's characterization of context is closer to the general description of mine. Cf.: "Whenever a sentence is said, it is said at some particular time, place, and world. The production of a token is located, both in physical space-time and in logical space. I call such a location a *context.*" (1980, 85)

conception of how the agent's commitment demarcates the context of a content: namely, the context of a content *is* its *context of use* (use in the sense of illocutionary performance).

There is an important caveat at this point: In this framework the illocutionary act does not determine its context arbitrarily—Although the illocutionary force renders its content context-sensitive—i.e., although, by performing the content in a token act, it makes the discursive evaluability and intelligibility of its content to be *concerning* a context (the context of use)—the force does not state anything *about* its context.[38] In other words, the context does not alter the content, otherwise it will violate Frege's Insight that the assertoric force does not contribute to the content—The assertion is not a characteristic function and the context is not a parameter.

The consequence of this caveat is that the context of an assertion is *not* freely *shiftable*—to evaluate or discourse on the meaning of an assertion we always have to put its locutionary content into its context of use—As Lewis pointed out: "[admittedly,] we could speak a language in which 'As for you, I am hungry.' is true iff 'I am hungry.' is true when the role of speaker is shifted from me to you—in other words, iff you are hungry. We could—but we don't."[39] To use Lewis's terms, in such 'shifty' cases it is rather the 'index'—some features of the context—that has been shifted away from the index of the context, but not the whole context itself. And such case "does not figure in any rules relating truth of one sentence in context to truth of a second sentence."[40]

In the next section (§5.5), I will have more discussion on context-shift and how it in fact works on the dissociation of the condition of evaluability from the context. For now, the formulation (*C**) of an assertion's evaluability condition can be made to fit the Lewisian *contextualism*.[41] Contextualism is correct in maintaining that evaluation of an utterance is impossible without putting the utterer's original intention into consideration. It also rejects minimalism, because it holds that in evaluating a sentence we cannot simply take its 'literal' meaning at face value (mathematical calculations

[38] Cf. MacFarlane (2014, 85ff.). Cf. also Recanati (2010, 211f.).

[39] Lewis (1980) 84f.

[40] Ibd.

[41] E.g. cf. Lewis: "sentence *s* is true at context *c* iff *s* is true at *c* at the index of the context *c*." (1980, 88). Cf. also MacFarlane (2009). Henceforth, if without specification I will use the notion 'contextualism' exclusively in this Lewisian sense.

aside) but must always take in considerations of the context in which the sentence—qua locutionary content—is uttered. Although contextualism usually makes no claims about agent's commitment, it nevertheless largely agrees with my top-down pragmatic view that in comprehending or evaluating a thought it is logically or explanatorily primary to have the full-fledged Austinian assertion—which entails the agent's context of use—in view.

Facing a sentence or a statement, one is free to say yes, or no, or not sure. But in making these evaluations, we are not just classifying the extension of the statement according to some arbitrary standard—Imagine that John said "Climate change is a hoax." and I replied "No, it is not true." In this case, it is uninteresting how I report my classification of the statement as true or false, just like others would do the same or otherwise. Also, John will not understand my reply as a simple report of classification carried out on my behalf. I am not expressing an irrelevant idea or playing crossword, and in normal cases it does not suffice to simply say "No, *period*." or "end of conversation." Rather, John understands that I am disagreeing with him; and I am expected to give further reasons explaining why it is incorrect to assert, in *John's* context, that climate is a hoax.

However, we see that there is a difficulty for (Lewisian) contextualism in accounting for belief-comparison and disagreements: For me to evaluate John's assertion, I am also making an assertion on my own—because my evaluation would be unintelligible if I do not assert it. But since the contextualism forbids a truth-evaluable sentence from shifting its context, there is no way to understand the evaluability condition of my own assertion (which evaluates John's assertion) other than an iteration of the context for John to make his assertion. If we call the context of my assertion *the context of evaluation*, and the context of John's assertion *the context of use*, it seems unavoidable for contextualists to claim, ironically, that the context of evaluation plays no significant role in evaluating an assertion, but only the assertion's context of use does.

To make it more intuitive, we can illustrate the contextualist picture using a two-dimensional framework. (Let us for the moment technically allow the context to be seen as a parameter of the content's truth.[42]) Suppose there are three contexts: c_1, c_2, and c_3

[42] In later passages of the dissertation (in §5.4.3 and in §§6.2–6.3), however, I will show that the two-dimensional semantic picture, based on the parameterization through contexts, has serious

possible for making assertions with the locutionary content "Climate change is a hoax." and there are three agents with different epistemic standards making such assertion:[43] Agent 1 and Agent 2 assert at contexts c_1 and c_2, respectively, both agree that climate change is a hoax. Agent 3 assert at context c_3, which disagrees that climate change is a hoax.[44] The truth of the sentence will then be evaluated with respect to the contexts roughly as follows:[45]

		EVALUATED		
		c_1	c_2	c_3
USED	c_1	T	T	T
	c_2	T	T	T
	c_3	F	F	F

Fig. 5.1.

It is obvious from the table above that the truth-values of the sentences are solely determined by their contexts of use, while contexts of evaluation only play a minimal role of reiteration. For example, although Agent 2 thinks that climate change is a hoax, it nevertheless concedes that climate change is not a hoax from the perspective of Agent 3 (in evaluating Agent 3's assertion). Under specific conditions, such concession could be fine.[46] But it will be solipsistic and hence cannot be satisfactory if we adopt it as *the* general framework for assertion-evaluation. For this reason, although the contextualism fits quite well with the Austinian, top-down pragmatic view of assertion, it lacks the crucial motivation to explain disagreement and thus does not do justice to the assertion's evaluability.

---

difficulties, because it cannot properly account for the dynamics of norm-internalization and knowledge-growth.

[43] Or, it could be also a same agent who makes different assertions about the content at different times t_1, t_2, and t_3.

[44] Nothing forbids an agent from making an assertion with the ostensible locutionary content that she takes to be false. For example, Agent 3 could utter the sentence "Climate change is a hoax" sarcastically, or counterfactually, etc.—that is why for an assertion the context is important because there is no component in a singular sentence that can specify the assertoric force of the content.

[45] Here I adopt the same table from MacFarlane's discussion on contextualism in idem (2014, 106). Figure 5.2 in §5.5.1 is also adopted from the same passage of MacFarlane's.

[46] Such concession exemplifies a faultless disagreement, which I will discuss in §5.5.2.

5.4.2 The Social Constructivist View

To rectify the contextualists' problem, there must be a mutual accountability between the pragmatics of use and the pragmatics of evaluation: On the one hand, as the contextualism correctly maintains, we can only evaluate an assertion with respect to the assertion's context of use; yet on the other hand, the assertion must be held accountable for its discursive propriety not only in the context of use but also in the context that is qualified to evaluate the assertion without performing it. This latter aspect of accountability is missing in the contextualist approach. In this subsection I shall sketch what I call the *social constructivist* approach to the problem of accountability / evaluability, which stands in obvious contrast to the contextualist one.[47] Robert Brandom and John MacFarlane are two major representatives of the social constructivist view. In what follows I shall examine MacFarlane's suggestion that assertion is an act of commitment undertaking.

If we apply the same two-dimensional framework under which three agents interact with the locutionary content "Climate change is a hoax"—Agent 1 and Agent 2 assert at contexts c_1 and c_2 agreeing that climate change is a hoax, while Agent 3 asserts at context c_3 which disagrees that climate change is a hoax—then the truth of the sentence, as suggested by MacFarlane, will be determined solely by the epistemic standards in the respective contexts of evaluation (Fig. 5.2)[48]:

USED \ EVALUATED	c_1	c_2	c_3
c_1	T	T	F
c_2	T	T	F
c_3	T	T	F

Fig. 5.2.

Unlike contextualism, the truth-values of the assertions in this framework appear to be dependent on the contexts of evaluation, while the context of use plays an apparently

[47] MacFarlane has been portraying his view as a 'truth relativism'. But instead of the questions whether and how truth is relative—it might be, under certain specific interpretation, compatible with my view that the constitutive norms of assertion are pluralistic—I take its fundamentally intersubjective (thus social constructivist) mechanism of evaluability to be central to our discussion. I borrow the term 'social constructivism' from McDowell see idem (2008).

[48] This framework is proposed by MacFarlane see idem (2014, 106).

minimal role. However, an asserter's liability must be limited with regard to the context of use, and the evaluator will also need to be qualified by this limitation. Hence, we have to figure out what this limitation could be. One possible explanation would be that there is a higher-order context governing the qualification of contexts for evaluating assertions made at other contexts, but it would be obviously a regressive argument. Another possible approach, which is adopted by the social constructivists, is that it is qualified by the asserter's *commitment*.

Like me, the social constructivists emphasize that in accounting for the condition of evaluability of assertion, we must view the agent's act of assertion as making a commitment. But unlike me, who contend that the agent's commitment is inherently self-attributing, i.e. inherently doxastic, the social constructivists maintain that commitment is inherently intersubjective, i.e., it is motivated by the responsiveness to the attitudes of other agents across contexts.

For MacFarlane, there are multiple aspects of an agent's commitment in assertion, which include:

> (W) Commitment to withdraw the assertion if and when it is shown to have been untrue.
> (J) Commitment to justify the assertion (provide grounds for its truth) if and when it is appropriately challenged.
> (R) Commitment to be held responsible if someone else acts on or reasons from what is asserted, and it proves to have been untrue.[49]

According to such characterizations, an agent's commitment is a socially and intersubjectively well-established normative status. But my main complaint about this view is that it sets a too weak constraint on the cognitive conditions of the agent in making an assertion. Admittedly, (J) predicts that the agent is committed to justify her assertion if challenged. But being committed to (i.e. promise to) justify an assertion is not yet justifying it. For me, since the discursive commitment is inherently self-attributing, for an agent to be committed it must already imply that the agent has been challenging herself to the justification of the propriety of her assertion—In committing oneself to the discursive propriety of a content one presents the content as justified, not merely to-be-justified. Making a discursive commitment should be an essentially

[49] MacFarlane (2005, 318), cf. also idem (2011). I will have more discussion on the assertion-withdrawal (W) in §5.5.2.

painstaking act, not a preliminary or tentative one. But according to the social constructivists, it seems that to undertake a commitment one needs to do nothing more than tentatively putting forward a locutionary content at face value. It then appears to be suspiciously easy for the utterer to be qualified as 'occupying a node' of the normative network of reasons.

In MacFarlane's treatment, an assertion's context of use determines the *world* of evaluation, while the context of evaluation (what he calls the 'context of assessment') determines the evaluative *standards*[50]—Regardless what kind of explanatory role does the 'world' of evaluation play here, it is at least clear, however, that the world could be anything in the context of use *except* its evaluative standard.[51] In other words, for social constructivists the context of use is inherently an object of evaluation but not a 'subject'—The asserter is allowed to make a diagonal self-evaluation, but this is only supplementary and not essential.

According to my own approach, quite the contrary, the agent's commitment must be inherently self-attributing. It means that the asserter's self-evaluation is not a supplementary, extra feature—It is rather the insincere or abused assertions, not the sincere or genuine ones, that need extra treatment—In primary cases, asserting p and asserting the discursive propriety of p are not two separate acts. This is because the evaluability of the consequences of speech acts is only half of the story; The other half, namely the evaluability of the acts' performances, is equally important. Generally speaking, I think if there is no accountable evaluability condition available for performing an assertion—condition that is proposition-constitutive rather than merely post-propositional—there would be no evaluability condition conceivable for assertion-assessment, either.

So far, I have briefly examined two approaches to the problem of evaluability concerning contextuality: contextualism and social constructivism. We see that although both approaches deny the free shiftability of the context in evaluating an assertion, their

---

[50] Cf. MacFarlane: "A sentence S is true as used at a context c_1 and assessed from a context c_2 iff S is true at c_1,$\langle w_{c1}, s_{c2}\rangle$, where w_{c1} is the world of c_1 and s_{c2} is the aesthetic standard of the agent of c_2." (2014, 67)

[51] We cannot say, like Kaplan did, that the world in the context of use *is* the world in which "The Americas is the Indies". If the content uttered in a context can already fix its descriptive terms within a definitive world like this, the utterance would be always true and an evaluation of it would be superfluous. For more discussion cf. the next subsection.

denials are based on different assumptions: For contextualists, denying shiftability is to emphasis the dominant role of the context of use (and therefore they reject minimalism); But for social constructivists, denying shiftability is to trivialize the normative significance of the context of use—For them, there is no profound (i.e. propositionally significant) pragmatic mechanism that serves to demarcate an assertion from other acts—For example, in Figure 5.2 when the sentence "climate change is a hoax" is uttered at c_3 insincerely, it is nevertheless taken at face value and hence evaluated at c_1 and c_2 as true. Hence, in the social constructivist framework there is no requirement for the agent's sincerity. Viewing assertions in this way thus allows the commitment to be—in Austin's words—'abused' or 'purported'.[52]

The social constructivists might reply that they are not trivializing the normative significance of the context of use, but 'socializing' or 'institutionalizing' it—They might contend that the evaluability of a speech act should always be explained in terms of its 'downstream' evaluation of its consequences, while the evaluability of *that* act of evaluation is explained in terms of the further 'downstream' evaluation of *its* consequences… and so on—They maintain that this is not a chain of regress but a network of society. But to me it is unclear how this 'network'-argument could be different from a Quinean one—In a causal history of ours, it might be that the self-evaluating performance is a latecomer to the networking of norms, but for our (philosophical) understanding of meaning and discursivity, I think self-evaluation should be explanatorily primary.[53] To conclude, I claim that it is not just an interest in

---

[52] E.g. MacFarlane: "One can lie without violating the constitutive norms of assertion." In idem (2003, 335).

[53] This also constitutes a typical objection to Brandom's inferentialist framework. Cf. McDowell (2005b), (2010), and Rödl (2000). For Brandom, there is a fundamental distinction between language–language moves and language–entry moves—we make observational reports in this latter sense, which are non-inferential and not intrinsically normative, but enabled by only having reliable differential responsive dispositions. (Cf. Brandom 1994, 219f.) However, such a non-inferential report does not qualify as a discursive act. When committing myself to asserting such a report, I would be only proposing or holding out a raw material for evaluation (scorekeeping) without having any condition of propriety constitutive in my act, i.e. I only interpret my observation behavioristically due to previous training through acquiring multiple reliable differential responsive dispositions, which is not different from a psychological mechanism.

In fact, we can imagine a whole colony of Martians (or, perhaps, an artificial neural network with reinforcement learning ability) making non-inferential reports with complex mechanism of dispositional attitudes and attitude-attributions, but we do not yet expect any

theoretical comprehensiveness to incorporate sincerity and self-evaluation into the framework of assertion; quite the contrary, it is the foundation of assertion.

I think the constitutivism proposing such a view neither needs to be captured by the skeptic problems as K-constitutivism does, nor needs to be captured by the solipsism as the Lewisian contextualism does—We can view the doxastic commitment as expressing a form of 'self-trust' and differentiate it from the problematic 'self-reliance'—This distinction is made by Linda Zagzebski in her (2012): She contends that self-reliance speaks of the idea of pre-given intellectual autonomy, which is presupposed in modern society as a form of what she calls an "epistemic anarchism" (2012, 7). Advocates of self-reliance typically argue as follows: "I trust reasoning; therefore, I trust my own reasoning. I trust conscientiousness; therefore, I trust my own conscientiousness. I trust human perceptual powers; therefore, I trust my own perceptual powers." (2012, 209) However, we cannot derive the reliability of our first-person thought from the reliability of a certain kind of knowledge of ours; quite the contrary, it goes in the reversed order: "I trust myself in particular when I am a certain way [i.e. being conscientious], and I see when I am conscientious that trust in my own possession of that property commits me to trust the instantiation of the same property in other persons. I trust the quality of conscientiousness because it is a rational commitment of self-trust. Self-trust is prior to trust in a quality such as conscientiousness or a faculty such as reason. It is prior to trust in any property at all." (ibd.)

When I assert that "I am YZ" or "I have leg pain", I thereby refer to the capacities and know-hows of perceiving, discerning, reporting, memorizing, proprioception, and bodily awareness, etc. and express trust in these capacities. But it is only possible when they are represented by my discursive power as constitutive of my

---

discursive act or conceptual content to be derived from such reports—If this characterization is correct, it is then unclear how anyone, qua reporter, can get a grip on the social inferential nexus of language–language moves. (The late Brandom seems to be aware of this problem cf. Part Three of his manuscript *A Spirit of Trust*.)

The difference between my approach and the social constructivists', therefore, is that such kind of non-inferential (hence propositionally non-constitutive) normative mechanism is not central to my framework of the pragmatics of thinking and asserting. It is true that in terms of a causal process we start to learn by being taught from external discipline. But as I will discuss later, this is only half of the story. We cannot understand the nature of learning and education purely in this way.

commitment to the discursive appropriateness of my assertions as following the relevant norms in a form of self-trust. Of course, disagreements and mistakes could arise in these assertions, but it is not, as social constructivists assume, a "conflict between self-trust and trust in others", rather, it is a "conflict that arises within self-trust".[54]

More work needs to be done to defend this point. After taking a closer look in the following subsection at the two-dimensional framework underlying the contextualism and the social constructivism, I will discuss issues about disagreement and mistake in §5.5.

5.4.3 The Two-Dimensional Framework and the Common Ground

We saw that in accounting for the pragmatic significance of context-sensitivity of the committal act of assertion-making, constitutivism encounters a similar puzzle like Frege did. Frege's Puzzle shows that it is difficult for an account of assertion to accommodate two things: First, the conceptual content of an assertion, given its context, should be *semantically stable*; And second, the assertion of an empirical content must be *fallible*. The requirement of semantic stability cannot be satisfied if we adopt a contextualism that takes contexts to be freely shiftable, because according to such kind of contextualism there is no way for the evaluations of an assertion to be misunderstanding-proof. The second requirement, on the other hand, cannot be satisfied if we adopt the Kaplanian approach to rigidified descriptions, since it, as indicated earlier, renders the act of assertion rather trivial.

Frege's Puzzle reflects the difficulty in handling the apparent tension between interpreting a sentence regarding its context and evaluating its truth—It seems that the condition of understanding a proposition and the condition of evaluating it are separate: I might have misunderstood someone and thus falsely interpreted an utterance, but nevertheless made a correct evaluation of the content; Or, I might have correctly interpreted an utterance but falsely evaluated its content—And this is indeed what originally motivated the application of the two-dimensional framework by Stalnaker.

However, as Stalnaker himself stresses, his motivation of introducing the two-dimensional framework is *precisely not* to hold understandability and evaluability apart;

[54] Zagzebski (2012, 210).

quite the contrary, it is to unify them into a singular theory of the singular evaluative act of assertion-(re)interpretation.

For Kaplan, since interpreting a character and asserting a content are two steps, we can take the two-dimensional framework at face value: It is one thing to correctly (or incorrectly) interpret an utterance, and another to evaluate it as true (or false). But the utterance-interpretation, therefore, works only at sub-sentential (or 'sub-propositional') level—For example, it concerns what the utterer is really referring to by using the words 'Phosphorus', or 'Tully', or 'physician', etc.—It is only about saturating a concept by letting one or another object falling under it, but not about the conceptual content or the proposition itself falling under one truth-value or another. I have argued in Chapter 2 that such an approach is captured by Frege's Puzzle and hence unsatisfactory. Contrasting such approach, Stalnaker points out correctly that the application of the two-dimensional framework is not really about laying out the different truth-values in a two-dimensional table (what Stalnaker calls a 'propositional concept'); rather, it is about the *transition* of the frameworks.[55] Only so can we establish a relation between the differentiating contexts between use and evaluation at a propositional level, not a 'sub-propositional' one.

For Stalnaker, the ideal transition for a propositional concept is an operation which he calls '*diagonalization*'. Its idea is to turn a two-dimensional propositional concept with any possible combination of differentiating contexts between use and evaluation into a standardized one—by projecting the diagonal cells onto the horizontal. In other words, for Stalnaker the standardized, diagonalized propositional concept takes the form as showed in Fig. 5.2.

Therefore, Stalnaker's view of the evaluability condition of assertions is similar to the social constructivists'. He does not fancy the idea that a standardized proposition might take the form like Fig. 5.1, which has its diagonal cells projected onto the vertical. Just like the social constructivists, for Stalnaker Fig. 5.1 only stands for the 'upstream', input proposition, which is rather an object expression *to be* evaluated. This object expression is then uttered and evaluated in a standardizing act—i.e. in a singular

[55] See Stalnaker (1978). Stalnaker further maintains that his two-dimensional framework is not used for semantics but 'metasemantics', and he takes the transition of the propositional concepts to be a metasemantic functional operation. But I will argue later that there are limitations in understanding the transition in this way.

act of (re)interpretation—as a downstream, output proposition in form of Fig. 5.2. This is how the operation of diagonalization or standardization works.

It is reasonable for Stalnaker to reject Fig. 5.1 as representing a viable candidate for standardized proposition. I argued that the Lewisian contextualism representing Fig. 5.1 is not eventually unsatisfactory. The Lewisian contextualism emphasizes that to evaluate an utterance in a semantically stable way—to avoid misunderstanding in interpreting that utterance—we need to refer backwards constantly and completely to the utterer's original context of use. But such a strategy is obviously expensive and time-consuming—The hearer must keep contextualizing the utterer's 'upstream' norms and intentions to make sure the utterer is understood correctly. And worse still, it ultimately leads to a 'maximization' of the context of use and ends up with solipsism, hence betrays the possibility of communication. The social constructivism, by contrast, is inexpensive and time-efficient. The hearer can largely ignore the original context of use and still make assessments based on the minimal, literal meaning conveyed by the utterer's locutionary content as long as they can carry on with the discourse further 'downstream'.

But I also argued in the last subsection that the social constructivism is equally unsatisfactory; namely, it goes to the other extreme that the context of use is minimalized, which ultimately sacrifices the cognitive significance of the utterer's act of commitment. So, what a Stalnakerian reply to my objection could be? The Stalnakerians would argue that the diagonalized propositions—propositions whose contexts of use are minimalized—are *ideal* consequences in communication: They agree with me that in reality both the context of use and the context of evaluation must be taken into account in communication. They call the context of use in a conversation its 'common ground' and maintain that a successful conversation should *reduce* its common ground. For example, in order to evaluate John's assertion about Phosphorus (to converse with him about it), I must have the conversation's common ground in view, which can be seen as a set of possible worlds in which John's use of the notion 'Phosphorus' may designate different objects. But for a successful conversation the possible worlds should be trimmed or eliminated. In ideal cases, therefore, I *no longer* need to contextualize John's assertion, and his use of 'Phosphorus' can be interpreted literally *salva veritate*, so that the two-dimensional framework is diagonalized into the 'downstream', output proposition—We tend to share information and give reasons in a

way with the minimum need of contextualization, because this is the most efficient and economical way to communicate.[56]

But I think Stalnaker's treatment of common ground has serious difficulties. Above all, the idea of common ground reduction relies heavily upon the metaphysical idea of shrinking the set of possible worlds, namely, in asserting we add more contents to the presupposed common ground, which, in effect, reduces the common ground—For Stalnaker, the only way for us to increase knowledge or information is to reduce possibilities. But this is puzzling. Intuitively, not all cases of our knowledge-growth are based on elimination of possible worlds. The inferences we make are not always monotonic, but usually ampliative. Later in §5.6, I will show that this problem is similar to the problem exhibited in Plato's *Meno*, which makes the activity of learning new things hard to explain.

I am sympathetic to Stalnaker's view that our discourse or conversation is not just about reiterating an utterance in a *context-preserving* manner, rather, it is inherently about evaluating the utterance in an act of (re)interpretation in a *context-changing* way. Call such a view the *dynamic view* of assertions and their contexts. Both the Stalnakerians and the social constructivists attempt to account for such a dynamic view. But my dissatisfaction with both approaches is that they failed to address the cognitive significance of the committal act of assertion, hence their attempts to account for a dynamic view are problematic—For Stalnakerians the change of context is a metasemantic, functional operation on the common ground, while the applicability of such an operation is still captured by Frege's Puzzle. (This time it is unclear how to coordinate its metasemantic stability and potential impropriety.) Moreover, it is worth pointing out again that the crux of the problem for the two-dimensional semantics here is that it failed to realize that the committal act of assertion cannot be seen as a function relativized by the context-parameter, according to which the assertion will hence be always evaluated at a shifted context. The social constructivists, in turn, are more sophisticated in this respect, since they do not insist that our inferential discourse must be monotonic; instead, they allow cases in which the committed utterer *withdraws* his

[56] MacFarlane also explicitly supports this idea. In chap. 12 of idem (2014), he argues from an engineering perspective that backwards contextualizing utterances is expensive and consumes too many intellectual resources. Keeping track of history is indeed burdensome—it is not only a burden of memory, but also in the sense that our history seems never stabilized. But I think this is what we have to live with as finite intellects. For more discussion see my §6.2.4.

assertion. Nevertheless, I shall argue later in this chapter that the idea of assertion-withdrawal is also problematic and cannot properly account for the dynamic view. Finally, I will propose my positive account of the dynamic view in Chapter 6.

The nature of assertion is not to simply share pieces of information, but to converse agreements and disagreements. And according to the dynamic view, to understand this discursive nature of assertions we need a proper understanding of the context-change in evaluating an assertion. In what follows, I will further facilitate a discussion on some controversies about context-shift and disagreement, to show that for a proper understanding of assertion a dynamic view must be possible.

5.5 Disagreement

5.5.1 Context-Shift vs. Context-Revision

I agree with the social constructivists (and also Stalnakerians) that the interpretation of an utterance constitutively involves evaluating it, which could yield either agreement or disagreement with the utterer—In order to agree or disagree with someone there must be more than one context at play; moreover, the agreement or disagreement is represented inherently through context-change rather than context-preservation. But the term 'context-change' is potentially ambiguous. Namely, a 'change' in context might sometimes be understood as a *context-shift*, which, however, is *not* a standard case of context-change. Therefore, we may call the standard case of context-change *context-revision*, to differentiate it from context-shift, which, I shall argue, rather represents the case of context-preservation. This subsection is chiefly about context-shift.

When an agent shifts a context away from her 'current' context of discourse, she thereby creates what Stalnaker calls a 'subordinate context'. Sentences in a shifted context are insulated and therefore exempt from evaluations external to that context. There are many uses of speaking in shifted contexts. Most prominently is Frege's example of fiction: When an actor recites his lines on stage, the propriety of his speeches is only relative to the context of the play, and it thus intrinsically makes no sense to agree or disagree with his lines from the perspective of the real world since the context has been shifted.

Context-shift should not be confused with world-change. World-change is speaking with supposition of another possibility, usually counterfactually or

subjunctively. Consider the following example, when a billionaire speaks to his children:

(12) If I am dead, my children, you are rich.

By making such statement, the utterer is aware, and also intends his hearers to be aware, that the world or the evaluability condition of his speech has been changed or qualified. The qualification is implicit in the presentation of this thought. Hence, we can utter the same thought in a subjunctive mood:

(12*) If I was dead, my children, you would be rich.

Even though these two sentences express a fictional situation, they are not speaking in a shifted context. In other words, though the content of the speech is partially fictional, the speech act itself is not fictional but committed to the evaluative standards in reality. By contrast, consider the billionaire speaks in a recorded message to his children:[57]

(13) I am dead, my children, and you are rich.

This is a case of context-shift. The fictional characteristic of (13) is not about any specific mode of presentation. Instead, it is a case of which the context of the whole utterance is shifted from reality. Similarly, what Recanati calls a 'direct speech report' is also a kind of context-shift:[58]

(14) … and then John said, 'I'm fed up with all this!' And he walked away.

The reported speech in (14) is neither a committal direct speech nor an indirect speech, but a speech displayed by the speaker locutionarily (i.e. phatically and phonetically) but not illocutionarily. Just like an actor on stage, in making this report the speaker of sentence (14) herself is neither agreeing nor disagreeing with the reported content.

In §4.2.2, I argued that there are two kinds of distancing act, i.e., two kinds of presenting a thought as non-committal: the qualified act of illocution and the disqualified act of illocution. Obviously, a speech displayed in a shifted context is a disqualified act—The context-shift is actually the disqualifying condition that renders the content pathological.

---

[57] I borrow this example from Recanati (2010, 198).

[58] See Recanati (2010, 194).

One might object that speeches in a shifted context is only exonerative but not pathological. But I do not think there is any real difference between the two. A pathological condition indicates that the utterer is bewildered or not self-conscious and therefore can be excused from being held accountable for his speech. The same goes to the actor who speaks on stage. It is sometimes claimed that a good actor acts in an ecstasy—He acts as if he is no longer himself—Or we might lose ourselves in an attractive novel or film, etc.—It must be a terribly bad novel if it allows its readers keep reminding themselves that there is a context identical with their own in which the writer sat in front of his writing desk and typed those sentences out.

Beside the examples above, there is still another important example of context-shift, namely a mishap that surprises the agent—The fact that it is a mishap instead of a genuine mistake indicates that the agent is exonerated from the unforeseen consequence, and one is exonerated precisely because he was pathological and ignorant in respect of that consequence. In such case, though we can say that the consequence does not agree with the agent, it does not constitute a genuine disagreement, but only a faultless one—I will come back to this later.

In these cases, we see that it is impossible to interpret an utterance without locating it in its context of use, but for the same reason it makes no sense to agree or disagree with the utterer in a shifted context, i.e. from outside. Therefore, the 'interpretation' of an utterance in a shifted context is not evaluation but rather only locutionary iteration of it, and we do so in a context-preserving way, i.e. indicating that it has a shifted context.

In contrast to context-shift, I agree with Stalnaker that any evaluative act which agrees or genuinely disagrees with a content must involve context-change (or what I call the context-revision). Now suppose (let us, again, allow the context to be technically seen as a parameter of the content's truth for the moment) that there are three contexts: c_1, c_2, and c_3 possible for making assertions with the locutionary content "Climate change is a hoax". Agent 1 and Agent 2 assert the content affirmatively at contexts c_1 and c_2, respectively. Agent 3 assert at context c_3, which disagrees that climate change is a hoax. Moreover, suppose as for Agent 3, while c_1 is a shifted context, c_2 is not. We can draw a comparison between context-shift and context-revision as follows:

		ITERATED	
		c_1	c_3
USED	c_1	T	T
	c_3	F	F

Fig. 5.3
< Context Preservation >

		EVALUATED	
		c_2	c_3
USED	c_2	T	F
	c_3	T	F

Fig. 5.4
< Context Revision >

It is understandable that these two matrices correspond to Fig. 5.1 and Fig. 5.2, respectively—Just like contextualism, Figure 5.3 indicates the preservation of 'upstream' contexts (and the context-shift in non-diagonal cells); but in such cases the content is only iterated or displayed rather than evaluated. Figure 5.4, in turn, indicates the 'downstream' revision of the context just like social constructivism.

Therefore, for Agent 3, we may say that he neither agrees or disagrees with Agent 1 because the latter is only, say, making a direct speech report. But he genuinely disagrees with Agent 2 at takes what Agent 2 utters to be false—Agent 3 evaluates the proposition according to his own epistemic standard implicit in c_3; moreover, his negative evaluation implies also *the denial of the validity or authority of Agent 2's epistemic standard implicit in* c_2.

I would like to make two remarks at this point:

First, although we say that context-shift and context-revision are two different kinds of content-interpretation, the context-shift is only explanatorily secondary. Locally speaking, every sentence in a shifted context is insulated from external evaluation since the illocutionary force is shifted away. In other words, the evaluative access across shifted contexts is 'blocked' or 'suspended'. But globally speaking, because there is no content without force, and because—as argued in Chapter 4—the locutionary act is only secondary, the disqualifying act of context-shift itself, as an act of distancing or suspension, is derivative of the act of illocutionary commitment.[59] Hence, Agent 3 needs to *do more* to determine that Agent 1 is exempt from evaluations—by acknowledging the context-shift—*not less*.

[59] Without the global background of shiftiness or relativity, we would be unable to tell that your judgment of taste is indeed a judgment of taste, or my recitation of the lines of a play is indeed recitation of the lines of a play—These locutionary acts would be sheer meaningless display of locutions.

Second, the comparison between Fig. 5.3 and 5.4 is only auxiliary—Both Stalnakerians and social constructivists employ such a comparison, but they nevertheless fail to account for a proper dynamic view, or so I argue—According to the dynamic view, context-revision is not only a product, but also a process; namely, it is rather *a transitory movement from preserving a context to revising it*.

But isn't it contradictory to the first remark that context-shift or context-preservation is only explanatorily secondary? The only possible explanation, I think, is that such transition is *recognitive*, rather than *explanatory*. Therefore, we have two opposite orders: the *order of recognition* from context-preservation to context-revision, and the *order of explanation* from context-revision to context-preservation. I will argue—in §5.6 and more explicitly in Chapter 6—that these two orders are two aspects of assertion-evaluation, which cannot be reduced to each other. But before that, I would like to say a bit more on the epistemic significance of distinguishing the genuine disagreements from the faultless ones.

5.5.2 Faultless vs. Genuine Disagreement

Faultless disagreement is not really disagreement but simple difference on contents. To say a difference is simple I mean the different contents in comparison are compatible with each other—As discussed in the last subsection, such difference implies a shift in context. We can compare this to the properties of objects: Object 1 is cubical, while Object 2 is spherical. We can say they 'disagree' with each other in terms of their geometrical shapes, but there is nothing incompatible here because two separate objects are involved.[60] Similarly, when two assertions faultlessly disagree with each other they are compatible, because they are just asserted in different contexts while the evaluative access across the contexts is suspended. On the contrary, a genuine disagreement indicates that the contents' difference is incompatible. When two assertions genuinely disagree with each other, it means they are evaluated according to a same standard or norm and at least one of them must be rejected. (Note, furthermore, that just like that context-shift cannot be independently understood but only in terms of derivation from

[60] We can also understand the logical secondariness of the faultless disagreement to the genuine disagreement through this analogy: namely, in order to grasp the difference between cubical and spherical it presupposes the concept that they would be incompatible if applied to a singular object.

context-revision, faultless disagreements correspondingly cannot be independently understood but only in terms of derivation from genuine disagreements.)

A common example for faultless disagreement is judgment about taste.[61] But it can be extended to all kinds of simple difference between the constitutive norms governing the agents' assertions. In §4.2.4 I made an example that Sophie, who was never trained in botany, and some botanists are asked to sort three trees into two groups. I argued that both Sophie and the botanists have performed appropriate discursive commitment since both of their acts of assertion followed their respective constitutive discursive norms and met their epistemic standards implicit in their contexts respectively. Hence their disagreement is an example of faultless disagreement.

Distinguishing genuine disagreements from faultless ones, therefore, means that we should distinguish *internal* from *external norms*. Internal norms are those that an agent has committed herself to follow in her act of assertion; external norms are norms that the agent has *not* committed herself to follow and thus is not authoritative and not constitutive of her assertion. It shows us—as I argued in Chapter 4—that the agent's discursive power plays an essential role in delineating the boundary of its scope of commitment—In making a commitment, the agent is conscious of her own discursive power as giving the constitutive unity to the normative reasons that she subjects herself to.

That certain norms are *not* constitutive of our agency is a central feature for finite intellect. We are all, in one way or another, ignorant. But here we can distinguish between at least two kinds of ignorance: On the one hand, there is ignorance that can be explained as certain unfavorable condition *in spite of* or *beyond* my power, either because I am pathological or because I am simply non-omniscient. And I do not hold accountability for the consequential errors or imperfections of my assertion if the relevant norms are beyond my know-how. Such kind of ignorance are disqualifying conditions for that respective discursive accountability—To use former U. S. Secretary of Defense Donald Rumsfeld's words, they are something 'unknowingly unknown' to the agent—When Sophie tried to sort the trees according to her everyday experience, the criteria for sorting American and European hornbeams were unknown to her in this sense—Yet on the other hand, there is also ignorance in the sense that one failed to respond appropriately despite the fact that it is within her power to do so. This is the

[61] Cf. e.g. Recanati (2007).

central meaning of what we call *mistakes* (or we might want to call it: 'genuine mistake'). By making a mistake, I not only *should have* but also *could have* avoided it, because the criteria for asserting appropriately were, in fact, 'unknowingly known' to me.[62]

In sum, one makes a genuine *mistake* if he fails to satisfy an internal norm, while, in contrast, if one fails to satisfy an external norm, we can say that a *mishap* occurs. A theory of discursive commitment should be able to distinguish between mishaps and mistakes and offer a balanced account for both—That is why I suggest that we need to take into account the agent's discursive power, for I believe it plays a useful role in properly understanding the distinction between internal and external norms—Unlike Kantian constitutivism, you cannot be committed to be an autonomous agent without having the power or know-how to follow relevant norms—If any norm that was not met can only be an external norm, as Kantian constitutivism in effect suggests, then all our failures and errors would only be mishaps, and we would be exculpated from any possible error of our assertions. But vice versa: if all our failures and errors are genuine mistakes, we would have an error theory of empirical judgment, which is equally problematic.

Social constructivists such as MacFarlane do not support error theory. They agree with me that the evaluative standard of an assertion is always sensitive to agents and contexts, and hence the constitutive norms for making assertions must be pluralistic—or as MacFarlane contends, relativistic. Nonetheless, I think the social constructivism fails to provide a balanced account of distinguishing between mishaps and mistakes.

According to the social constructivist framework, the distinction between mishaps and mistakes is normatively trivial and perhaps only 'pre-social'. But I think this would yield a too strong treatment of our sociality and accountability. If I tell you I will arrive in Munich on Sunday by air, my assertion should be understood with epistemic implicatures such as that it is only taken to be true under the condition that the

[62] Saying that there could be something 'unknowingly known' to me might sound self-contradictory according to our commitment-based account of agency. Indeed, as I will discuss in more detail in Chapter 6, this could only be expressed by the agent in terms of backtracking counterfactuals.

flight by then is still available.[63] If making an illocutionary commitment means to me that I thereby undertake an 'infinite responsibility' for any inferential consequence my assertion might have, I would be understandably deterred from performing any assertion at all.

The social constructivists could reply to me that I do not need to be specifically painstaking to undertake a commitment—I only need to *withdraw* or *retract* the assertion when it is shown to be false.[64] But I think the idea of assertion-retraction is confusing at this point: I do not need to retract or revise my previous assertion in the case of faultless disagreement, given that it is indeed a mishap instead of any avoidable error due to carelessness, etc. If an unexpected storm or any kind of 'force majeure' occurs so that I can no longer manage to arrive in Munich on Sunday, I will not need to abandon any of my previous beliefs about my flight but to simply make another one. We can compare this to another example: If I say "I don't know whether it will rain on next Sunday", it has the implicature that I am ignorant of the whether *until then*. So when the day comes and the weather becomes known, there is no need for me to retract or revise my previous assertion because it has expired and is hence faultless in this regard.[65]

Admittedly, when it comes to genuine disagreement we indeed need to revise our beliefs. I also agree with social constructivists that it must involve the denial of the validity of the epistemic standards implicit in the context of a mistaken assertion. But even in such cases of genuine disagreement, I think the idea of assertion-retraction is still not helpful—One difficulty is the so-called 'belief recovery' problem in belief dynamics (which is much debated in model theory and computer science): Namely, if retracting a belief involves what Stalnaker proposes the reduction or elimination of possible worlds, then it is unclear how we can recover a retracted belief since it seems impossible to add those already reduced worlds back; but if retracting beliefs does not

[63] Such conditions may be expressed as "to the best of my knowledge". But its content can only be grasped in contexts and implicatures, not as a proviso explicitly stated or strictly deducible from the asserted content. Cf. §4.2.3.

[64] See §5.4.2.

[65] Implicatures are not always about rendering the utterance faultless. E.g., if I claimed that John has three children but it turns out that he has four, the implicature that I have expressed the maximum number should lead one to recognize that my assertion is a mistake.

involve world-reduction, it is unclear what a retraction is and how it interacts with the agent's doxastic context or common ground.[66]

The difficulty to model the procedure of belief-revision aside, the major difference between social constructivism and my approach is that I do not think the act of genuinely agreeing or disagreeing with an assertion can be governed simply by external norms. Genuinely disagreeing with an assertion is an act of giving reasons rather than simply of coercion and punishment. Although the agency of discursive commitment is always socially constituted, that must not mean that the boundary of a commitment could be arbitrarily settled by an outside standard regardless of the agent's knowledge or power.[67]

An agent of mishap is pathological and hence *disqualified*, while an agent of mistake is *qualified*. Accounting for the qualified agency hence requires us to properly distinguish between the two. But since social constructivists dismiss the distinction between mishap and mistake, they dissociate the act of retracting an assertion from the act of admitting fault.[68] For them, just like there is no intrinsic relation can be established between commitment and sincerity, there is equally no way to intrinsically determine whether one who retracts an assertion is being forced or pretending to do so. Since the social constructivists explain context-sensitivity exclusively in terms of downstream evaluations, for them, as long as a downstream evaluation of genuine disagreement is reached, the permissibility of the upstream content can by no means be understood as intrinsically logical or normative. But it is then unclear how the recognitive order of progression from the upstream permissibility to the downstream non-permissibility could be intelligible.[69]

---

[66] A summary of relevant debates cf. Hansson (2017). More discussion on belief dynamics cf. §6.2.

[67] Was Columbus genuinely mistaken when he asserted that the land he reached is the Indies? Now we can see that it is a question that cannot be so easily settled. One cannot be genuinely mistaken of something he was completely ignorant of. But Columbus was not ignorant of navigation and geography… (And his assertion was relatively quickly falsified after his death.) Maybe we can say he was one of those in his time closest to be genuinely mistaken in this matter… Anyways this is obviously not a standard example for our discussion since it is more or less anachronistic.

[68] Cf. e.g. MacFarlane (2014, 110).

[69] Another example of overlooking the significance of the order of recognition is Strawson, who writes in his *Individuals*: "I have spoken […] of conceptual *steps* or *transitions*, as if I were

Finally, here we are having a difficulty similar to that in the debate of the 'double effect' of moral actions: Saying that an assertion has made a mistake is finding fault and hence genuinely disagreeing with it. Yet a genuine disagreement is irreducible either to a faultless one or to a simple affirmation of something new. In the next section, I will suggest a dynamic view on the problem. I will argue that as users of natural languages we can only understand genuine disagreement from inside, i.e. in terms of self-disagreement and hence in a sense in terms of *regret*[70]—For user A to genuinely disagree with user B, A must acknowledge that B is qualified as a discursive agent to eventually agree with A. It involves a transition of B from being disqualified to being qualified, which is an *internalization* of epistemic standards.

5.6 Constitution in Change: Towards a Dynamic View

In this chapter I examined the most prominent version of constitutivism, i.e., the Kantian constitutivism, and its underlying difficulty. I also discussed its major rival theory, i.e. the social constructivism. The moral of the discussion is that for a theory of discursive agency to account for Thesis (β) it must maintain a balanced view on the relation between faultless disagreements and genuine disagreements, or between mishaps and mistakes.

As a finite intellect, my illocutionary and doxastic commitment cannot fully determine or exhaust the inferential consequences of my assertion. Hence the social constructivists have pinpointed the weakness of K-constitutivism, namely that by defending agency it has ignored that assertion-evaluation is always socially complex. However, so too is the act of assertion-making—The social constructivism nonetheless falls short of accounting for the significance of the speaker's knowledge in assertion.

In a conversation, the discursive agents not only exchange sentences and literal information but also convey knowledge. Yet for the knowledge acquisition it is not

---

speaking of a development in time […] Perhaps there are such stages […] Perhaps there are not. I do not know and it does not matter. What is in question is not an order of temporal development, but an order of explanation." (1959, 209).

[70] Similar to Austin's use of the notion 'infelicitous', being 'regretful' here does not refer to a sheer emotion but rather a pragmatic feature constitutive of understanding genuine disagreement. More discussions on regret see §6.2.

enough to adopt a belief granted as true—Knowledge is considered to involve conceptual innovation and hence more valuable than the simple status of possessing certain true beliefs. But it is unclear how to account for the possibility of conceptual innovation in a social constructivist framework.[71]

L. J. Cohen has formulated the difficulty as follows: "In order to be able to recognize malapropisms", he said, we have to maintain the distinction between "the actual meaning of the word in a particular utterance" and "the meaning that the speaker intends it to have there", while "in order to maintain this distinction we need to suppose that somehow the actual meaning is socially sanctioned while the intended meaning is determined by the speaker on his own". However, when "a new word-meaning makes its first appearance, […] it seems already to be an actual, and not merely an intended, meaning".[72] The difficulty, in other words, is that in a social constructivist framework there is no resource to explain how it is possible for us to learn anything and improve our discursive power or knowledge.

The difficulty can also be shown in terms of the agent's constitutive norms. We may distinguish between three kinds of norms that might be relevant in evaluating an agent's content: (i) *known norms*, (ii) *knowingly unknown norms*, and (iii) *unknowingly unknown norms*. According to our discussion so far, (i) and (ii) can be seen as the qualifying conditions implicit in the agent's assertion, explicable in terms of the agent's pro tanto reasons in deliberation. But conditions as such do not introduce any essential features of the agent into the asserted content but some empirical properties that (implicitly) belong to the agent—They must not behave as if they make up an ineliminable indexical in the assertion—That means, since (i) and (ii) are seen as something eventually known by the speaker, they should by no means constitute any ineliminable condition for the truth-evaluation of the agent's assertion. We do not want any ineliminable indexical in our assertions, just like we do not want any persistent, 'shifty', context-parameter in our assertions (as the indexical contextualism and the two-dimensional semantics do), which makes assertion-evaluation across contexts eventually inexplicable.

However, we have also argued against the deflationist or the semantic minimalist approach to the assertion, which rejects the idea of agency as either trivial or

[71] This is usually called the *Meno* Problem, cf. Pritchard, Millar, and Haddock (2010).

[72] Cohen (1986, 222).

redundant. Indicated by the agent's socio-historically mediated discursive power, which is explicable in terms of the pluralistic and use-context-sensitive constitutive norms, the constitutive agency is rather essential for assertion-evaluation. It means that it is still true to maintain that the evaluative norms are relative to each agent's own context of use or background knowledge. In a discourse, the interlocutors must coordinate their background knowledge and hence the norms implicit in their interpretations of the content at stake. But the question is, if (i) and (ii) cannot be understood as making up any ineliminable condition of the assertion, how is the coordination of the agent's background knowledge a necessary procedure for evaluation?

To understand the essentially socio-historical nature of discursive agency we need to understand our epistemic finitude, which demands that we must take (iii) also into consideration. And it is exactly the infinite character of (iii), I think, that allures philosophers to embrace something like an ineliminable indexical in conceptual content. The dynamic view of constitutive agency that I would like to propose in what follows, therefore, is an attempt to incorporate (iii) into the picture of assertion-evaluation as distinct from both (i) and (ii) to account for our epistemic finitude and the possibility of knowledge-growth, without adopting an ineliminable indexical or shifty context-parameter in the interpretation of asserted contents.

Learning is essential for us finite intellects, and we learn from mistakes. In exercising our capacities there could always be norms that we fail to follow. When we were born, we were basically ignorant of all these norms, and we live by building up our knowledge and skills—hence: our discursive power—so that we will not fail to follow those norms in future. I argued in §4.3 that believing, just like asserting, is an active process—It is an ongoing act of discursive commitment and accountability, because in order to keep track of and properly evaluate one's discursive commitment we have to look into his engagement in the act of discourse over time.[73] In this manner, the agent's constitution itself should also be understood as constantly changing or evolving.

[73] Strictly speaking, understanding assertion in this way already expands it into the wider notion of discursive acts in general, which no longer has its focus on singular assertions but rather on acts of reasoning, inference, etc. (cf. Appendix 2). Moreover, it may be argued that time consciousness is constitutive of a finite intellect—If my agency is timeless, it is unclear how I could still hold my previous self accountable, since it means either that I am eternal (so no change in time or in thought possible) or that I do not have a personality at all. Cf. Rödl (2012, 76ff.).

To differentiate such view of constitutive agency from the Kantian one, we can call it a dynamic view of discursive agency. And I take the recognition of mistakes, which introduces qualifications into our knowledge and skills, to be crucial for understanding dynamic constitutive agency.[74]

In a sense, the exercise of my discursive power is externally constrained, but as McDowell puts it:

> [such constraint] is authorized from within the practice of thinking, by norms that are constitutive of the practice. […] It is not clear how something other than further elaboration of the idea of norms internal to thinking could cast more light on the source of this normativity. [… The idea] is not that we confer authority on the norms in an act of legislation that brings them into being as authoritative, but just that they are constitutive of the practice of thinking, an activity in which we realize potentialities that are our own. We do not have a choice whether to be thinkers, but that does not imply that in submitting to the norms that thinking is anyway subject to, we abdicate any genuine freedom. (McDowell 2005c, 105f.)

Unlike K-constitutivism, there is no such thing as constituting yourself according to *the* law of self-legislation or self-constitution. But neither should we take the norms to be simply external—The practice of thinking or asserting is a social practice, but it should not be confused with law enforcement or coercion; Rather, it involves the internalization of norms, which signifies the course of learning. During learning, one's constitution of agency receives different determinate shapes, and thereby achieves the transition from pathology to knowledge.[75] Since our discursive power is in a constantly undergoing change, the autonomy in assertion is always in becoming. Learning, therefore, is to serve as the means to perfect our knowledge and power to achieve the

[74] Davidson characterizes the difficulty of understanding qualified agency as a difficulty of understanding irrationality, which, he famously notes, is a "failure within the house of reason"; and he concluded that to understand irrationality we must allow a degree of heteronomy to be partially incorporated in the autonomous agency. Davidson (1982, 169).

[75] McDowell suggests that the most accurate term describing this course is the German word '*Bildung*'. I agree. The idea derives from Hegelian philosophy, according to which the activity of the mind (*Geist*) is a course of emancipating conceptual development out from pathological settings (e.g. habit) through recognition and thinking into the ethical life (*Sittlichkeit*) of humans. Cf. also Rödl (2016).

end of autonomy. As Korsgaard vividly depicts, the act of constituting yourself as agent is a never-ending movement of pulling yourself together.[76]

Suppose Sophie sorted the three trees according to her perception of the colors of their leaves at c_1. She believed that the content of her assertion *p*, to the best of her knowledge, is discursively appropriate. I argued that Sophie was correct given that she had properly exercised her capacity of making observational report and hence was just faultlessly disagreeing with the botanists' way of sorting the trees otherwise. Now suppose that Sophie decided later that she should get botanical training. She recognized the botanical criteria for differentiating the trees into American and European hornbeams and asserted *q* at c_2, which excludes the appropriateness of *p*—She took her previous way of sorting the trees to be inaccurate—In the sense of context-preservation, *p* and *q* are just compatibly different because they are asserted in two different contexts with different constitutive norms. Yet in the sense of context-revision, Sophie believed not only that she had the knowledge to assert *q* at c_2 but also that she could have asserted *q* at c_1—She believed at c_2 that she in a sense had and in a sense had not the knowledge at c_1 to assert appropriately. In other words, at c_2, Sophie has two aspects to conceive of her agency at c_1: It is in a sense disqualified from and in a sense qualified for having the epistemic standards and know-hows to sort the trees with botanic accuracy. I think we can consider these two aspects as mutually irreducible and fundamentally inseparable for understanding the course of learning and change, which can be characterized as a process through genuine disagreement (by simply substituting 'speaker' for 'Sophie at c_2', and 'hearer' for 'Sophie at c_1').

In §5.5.2, we discussed the distinction between mishap and mistake, or correspondingly, between faultless and genuine disagreement. An important aspect of genuine disagreement, as argued, is that it aims at the hearer's acknowledgement of mistake. The hearer's acknowledgement of mistake aims at the hearer's recognition of a previously *unknownly unknown norm* as being *counterfactually constitutive* of his previous assertion (which thus has a unique and irreducible normative feature of backtracking 'regretfulness').[77] Now, according to the dynamic view, we can summarize the dynamics of disagreement as follows: *a)* In a discourse with genuine disagreement, the constitutive norms must undergo a dynamic process of change; *b)*

[76] Cf. Korsgaard (2009), *§6.4*.

[77] In Heidegger's words, in involves "knowing the unknown as known".

The process of change is explanatorily primary and cannot be reduced to differences between static phases (relativized by a shifty context-parameter, e.g. as in the two-dimensional semantics); *c)* The agent's background knowledge should not be represented by an ineliminable indexical, but neither is it only a trivial or reducible feature; rather, it is a feature of the dynamic process; *d)* In the case of knowledge-growth, the agent 'internalizes' certain norms, while such 'internalization' can be seen as a process of self-education that motivates disagreement-resolution and conceptual innovation. I shall make a more detailed discussion on this in the next chapter.

CHAPTER 6
Dynamics of the Discursive Power

Assertion, as argued, is an act of performative commitment to the appropriateness of the content at the context of conversation. It indicates that the agent's understanding of the content's appropriateness refers to the conformity to the epistemic standards and norms implicit in the background knowledge of the context. (In this respect, therefore, the question of demarcating meanings and truth-values is a question of demarcating the constitutive background norms.) Moreover, such an approach is a discursive one—We will need to conceive of the committal act as essentially consisting in the discursive movement of giving and asking for reasons. To convey meaning in a discourse is to give reason to the hearer, which, when successful, results in the hearer's integration of the given reason into her own beliefs.

The difficulty, however, is that the norms and epistemic standards related to a conversation are usually *not* symmetrical among its interlocutors. Hence, the major challenge to the approach of constitutive norms consists in the task of accounting for the asymmetry of norms in discourse—In such a discourse, the 'asymmetrical' norms are in a sense not implied, yet in a sense indeed implied, in the hearer's background knowledge—To account for this, I claimed that we need to have a 'dynamic' constitutive picture of the hearer's belief-integration as a process of norm-internalization. In this chapter, we want to take a closer look at this idea of norm-internalization and constitution-change.

In §6.1 I shall first give a general description of the acts of reason-giving, distinguishing them into three kinds. I will then proceed in §6.2 to discuss the mechanism of belief-integration as the consequence of reason-giving acts. I conclude by proposing a what I call 'historical revisionist' framework of belief-integration, according to which the discursive movement of giving reasons and integrating beliefs is taken to be conceptually primitive and irreducible. I will make a closing remark on the discursive power with respect to the dynamic framework in §6.3.

6.1 Three Kinds of Reason-Giving

The act of assertion, as the primary species of utterance, is the act of presenting and differentiating thoughts in discourse through agreements and disagreements. From this aspect, we might want to view assertions as acts of giving reasons. In this section, I would like to distinguish three kinds of acts of reason-giving. Corresponding to the three kinds of discursive agency distinguished in Chapter 4, we can equally distinguish the acts of reason-giving into the *unqualified*, the *qualified*, and the *disqualified* kinds. Just like any other speech acts, the act of reason-giving can have various modes of presentation, such as suggestion, declaration, instruction, etc. But our distinction between the three kinds does not directly concern the difference on those modes of presentation; rather, we should distinguish the three kinds in terms of the different ways in which the acts invoke agreements and disagreements.

6.1.1 Reason-Giving Acts without Qualification

The first, unqualified kind of reason-giving are conversations in which the interlocutors achieve maximum mutual intelligibility with minimum requirement of context-coordination. When the speaker makes an assertion in this way, she expects that the assertion will meet with no disagreement, confusion, doubt, or resistance of any kind at the hearer's side, so that the hearer can immediately accept the consequence and make the same assertion upon hearing without additional acts of asking for further reasons. In other words, the assertion's epistemic background and constitutive norms are supposed to be transparent to both the speaker and the hearer—In such case, the speaker can precede his sentence not only with "I think" but also with "We think". This kind of reason-giving represents the most efficient and frictionless way of information-sharing.

Note that this unqualified kind of reason-giving does not mean that we can share information in an absolute, divine way—A divine intellect, if conceivable at all, would be so perfect that it would never need any communicative act to gain information—Quite the contrary, as indicated in §4.2.4, for us finite intellects, to constitute oneself as agent without qualification means only that the qualification is (provisionally) absent. It is conceivable that after the speaker made an assertion in this unqualified manner, she might nonetheless—*to her surprise*—meet with the hearer's confusion or disagreement.

Surprise involves experience of dissonance in one's beliefs, which leads to recognition of disagreement.[1] Since an unqualified act of reason-giving should be free from disagreement, neither should the speaker's assertion surprise the hearer nor should the speaker be surprised by the hearer's reaction of it.

Here are some examples of such kind of unqualified act: Suppose that two people are walking together in the rain while one of them says: "It is raining." The setting of the conversation implies that the speaker knows that the hearer knows that it is raining, and the hearer also knows that the speaker knows that the hearer knows that it is raining… In fact, the epistemic standard and norms constitutive of the assertion are transparent to both the speaker and the hearer in this case. Hence it could fit as an example for unqualified mutual intelligibility. However, this example is not very ideal, since according to its setting the information seems perfectly symmetrical between the two interlocutors, and the speaker's communicative act seems rather superfluous—One can compare it to the case of the divine intellect—In such case, we would rather suspect that the speaker's speech is not yet complete and expect him to say more. E.g. he maybe wants to suggest that since "it is raining, we should to go a coffee shop instead of walking to the park." Or maybe the speaker uttered the sentence rather exclamatorily, alluding to his personal perception or emotion (e.g. annoyance) about the rain, etc.

But there are better examples. Suppose that John and Mary are in Berlin. John wants Mary to check the temperature in another city, say, Boston, for him. Mary then checks her phone and tells John: "It is 20." In making the statement—given that it is a successful and 'frictionless' information-sharing—Mary knows that John knows that Mary now knows that the current temperature in Boston is 20 degrees Celsius, while John indeed knows that Mary knows and also knows that Mary knows that John knows…—In sum, we can say that they* now know that the current temperature in Boston is 20 degrees Celsius, because it implies that the epistemic standard constitutive of the assertion (the reliability of the weather report on Mary's phone, e.g.) is transparent to both of them—Call this kind of collective transparency *mutual*

---

[1] Later in the subsections that follow, I shall argue that the experience of dissonance in one's beliefs does not necessarily leads to recognition of genuine disagreement and incompatibility; On the contrary, it might just lead to recognition of sheer contrast (and hence faultless disagreement or compatible difference)—This is typically the case when some breaking news or mishap occurs, which can be surprising but nevertheless does not contradict any of one's previous beliefs or expectations.

awareness—In the case of mutual awareness, therefore, Mary can precede her sentence with "We think" rather than "I think".[2] The explicit use of "I think", by contrast, might rather add a layer of indeterminacy to the content.

A major feature of this kind of reason-giving is that it requires minimum to zero change of the hearer's background beliefs except adding the new one. In other words, the procedure of knowledge-growth is basically monotonic—Since the hearer has no preexisting belief that is incompatible with the speaker's assertion, in believing in the speaker's assertion the hearer does not need to revise or knock out any of those preexisting beliefs. Admittedly, before being informed by Mary of the temperature in Boston, John was ignorant of it, and he knew that he was ignorant of it. But *that* belief of his (about his ignorance of the temperature) has the implicature that he was ignorant until he got informed by asking Mary to check it, and he expected this to happen. Therefore, when Mary later tells him the temperature *that* belief about his ignorance of the temperature simply expires rather than needs to be knocked out due to any incompatibility or falsehood.

6.1.2 Reason-Giving Acts with Qualification

The qualified kind of reason-giving involves conversations that are polemical or pedagogical in nature. It could concern persuasion, or instruction, or rejection, etc. Generally, the interlocutors in such conversations do not have the ideal mutual intelligibility and agreement but ask for further grounds or reasons to coordinate their contexts. They expect their conversation to be frictional and yield disagreements. Therefore, the epistemic background and norms constitutive of the speaker's assertion are supposed to be not fully transparent to the hearer.

---

[2] Even though we don't usually do this, it is certainly possible—especially when a third party is involved in the conversation (e.g. imagine that Mary and John are pilots making their flight plan, when Mary talks about the temperature in their destination city, she is aware that others, say, the air traffic control, is also listening on the frequency). But this exactly shows that although the unqualified committal act of speech is the central case of assertion, we still need to understand the necessity of performing such an act, which is only possible in a larger context of reason-giving discourse, which, in turn, always expects more or less epistemic frictions in the hearer and hence the socio-historical qualifications of discursive agency (e.g. the supposition that the air traffic controller, who also takes part in the conversation, does not know the temperature in the destination city).

The disagreements involved in qualified reason-giving are genuine disagreements. For example, when a botanist tells Sophie that she should rather have sorted the trees in a way that differentiates between American and European hornbeam, the botanist is implying that other ways of sorting the trees are inaccurate and thus mistaken. Since giving reasons in this way is to point out a mistake, the speaker takes the reason he is giving to be one that is not fully expected by the hearer. He intends that his assertion will generate friction and invoke dissonance in the hearer. In this sense, therefore, the speaker gives the reason from outside—i.e. to the hearer's surprise.

However, as I have argued in §5.5–5.6, this is only half of the story. Under the framework of dynamic constitutive agency we should rather comprehend such act of reason-giving as an internalizing act of learning. Namely, while in the order of recognition the speaker gives the reason primarily from outside, in the order of explanation, by contrast, the genuine disagreement can only be understood from inside, i.e. in terms of self-disagreement—For the botanist to genuinely disagree with Sophie, the botanist must presuppose that Sophie is qualified as a discursive agent to eventually agree with him and hence disagree with her own assertion.

To understand the dynamic framework of learning, therefore, the externality and the internality of the norms are two mutually irreducible and fundamentally inseparable aspects. It means that the botanist believes both that Sophie was ignorant of (and thus disqualified from following) the relevant botanical norms *and* that she is qualified to make assertions with the botanical norms being authoritative.

Two remarks need to be made at this point: First, the act of reason-giving does not necessarily need to involve multiple persons. Just as an agent can agree or disagree with herself, she can also give reasons to herself in a deliberative or retrospective manner. Both the *interpersonal* and the *intrapersonal* difference share the similar structure of self-disagreement in terms of assertion and reason-giving.

Second, a genuine disagreement does not necessarily concern explicit denial or rejection. It can also concern an attempt to simply convince the hearer, insofar that the act of convincing implies the speaker's awareness of potential friction from the hearer. The speaker does not expect that the hearer will internalize the reason in an unqualified way; instead, the speaker intends the hearer to integrate the assertion to her background beliefs while extracting or revising the preexisting beliefs that are incompatible with the assertion. Thus, the inferential process involved in such kind of reason-giving is usually non-monotonic and ampliative. I shall say more on this in §6.2.

To sum up, the speaker of a qualified reason-giving act intends the hearer to internalize the relevant reason or norm and hence be regretful. The speaker reasonably has such kind of intention when the setting of the conversation is in favor of a pedagogical or polemical discourse—When Sophie is a botanic student in a field trip, for instance—In such cases, there is an asymmetrical relation between the speaker and the hearer: What is asymmetrical is not only the information or the reason that is being given by the speaker to the hearer, but also the awareness that the reason is authoritative. Nevertheless, the intelligibility of this asymmetry presupposes the speaker's and the hearer's mutual awareness of the polemical or pedagogical setting of the conversation[3]—If, instead, Sophie refuses or fails to acknowledge the setting of the conversation, she will not recognize the relevant botanical norm as authoritative, and the speaker's act of reason-giving will also fail or misfire—Realizing such failure is to realize that the act was disqualified from being inherently reason-giving.

6.1.3 Disqualified Reason-Giving

In a disqualified act of reason-giving, the speaker does not intend or expect the hearer to recognize the reason or the norm he is giving to be authoritative. Hence strictly speaking it is not an act of giving reasons: On one hand, if the speaker wants to enforce the norm, it will be an act of coercion, i.e. an act that forces the hearer through order or threat to behave as if he follows the norm. In this case, the speaker is not treating his interlocutor as a person but as a sheer object—On the other hand, if the speaker does not want to enforce the norm, it is then an act of displaying a locutionary content in a shifted context. In such case, the speaker is invoking a faultless disagreement.[4]

Even though they seem quite different, both coercion and faultless disagreement are acts disqualified from being intrinsically meaningful. They are only meaningful in a more global context that explains the context-shift (the intervention or exoneration, etc.) Take the tree-sorting example again: If Sophie has never been trained in botany and does not have any interest nor obligation to receive any botanic training, it would be

[3] It can be argued that this kind of mutual awareness of internal difference or asymmetry indicates precisely the second person standpoint or the I-You-consciousness, which holds the agents apart in a joint action of self-knowledge. Cf. Darwall (2006), Haase (2014), Rödl (2014). Cf. also Appendix 3.

[4] Discussions on context-shift and faultless disagreement see §5.5.

intrinsically senseless to tell her that one should sort the trees into American and European hornbeams instead of according to their sizes—It would be a sheer display of locutionary act—Admittedly, it could be meaningful in a global context that explains the botanists' point of view with context-shift. But since the botanist's utterance is only meaningful at a shifted context, the botanic standard or norm for sorting the trees are not authoritative for Sophie, hence the utterance is disqualified from being reason-giving.

How does the belief-integration work for disqualified reason-giving? We have already seen that for unqualified reason-giving it involves minimum friction and revision of preexisting beliefs and hence is monotonic; and for qualified reason-giving it is non-monotonic and requires revision or extraction of incompatible preexisting beliefs. For disqualified reason-giving that expresses faultless disagreements, since the speaker does not intend the hearer to recognize the given reason to be authoritative, he certainly does not expect friction and belief-revision at the hearer's side. Instead, he expects the hearer to set up a subordinate shifted context and update information in that context.

Although a disqualified reason does not cause friction, it can still invoke the hearer's experience of dissonance and hence faultless disagreement, because it is something not expected by the hearer—to the hearer's surprise—though it is compatible with the hearer's background beliefs.

6.2 The Dynamics of Belief Integration

With the distinctions in the previous section about reason-giving, we are ready to give a more general discussion of 'reason-receiving', i.e., belief integration. I shall begin in §6.2.1 with a brief introduction to the belief dynamics and some of its basic notions in literature. To account for the process of belief-integration we need to conceive of some basic operations that operates on one's existing beliefs. I will introduce the basic operations needed for belief-integration—above all, the operation of belief revision—in §6.2.2. The notions of these operations are borrowed from the theories of belief dynamics in literature without adopting their formalizations. Instead, I will only give a discursive illustration of the dynamics of belief-revision—After discussing in §6.2.3

some basic feature of the counterfactuals at play, I will lay out my favored historical revisionist picture in §6.2.4.

6.2.1 Some Preliminaries on Belief Dynamics

Since the middle of the 1980s, especially in the wake of the so-called AGM model (named after Carlos Alchourrón, Peter Gärdenfors, and David Makinson)[5] formalizing the process of revisionary operations upon one's background beliefs, the logical theory for *belief dynamics* has gained much development—with possible applications in computer science e.g. database engineering. In my discussion on belief integration, I will not discuss the axiomatic formalizations of the theory of belief dynamics nor apply its calculative mechanisms. Nonetheless, I would like to use in my own investigation (§6.2.2) a few notions and symbols of the theory. Hence in this subsection I would like to make some (very brief) clarifications and comments on the philosophical background of the theory—I mainly consider two issues here: First, I will make a note on the distinction between *belief set* and *belief base*; and second, I shall say something on the relation between the operation of *belief revision* and the operation of *belief contraction*.

A *belief set* is a deductively closed set of sentences that an agent is committed to believe. Belief set theorists usually use a consequence operator 'Cn' to denote the function from a set of sentences A to the set of sentences $\mathrm{Cn}(A)$ that are the logical consequences of A. Since a belief set is deductively closed, for an agent who has a belief set as its doxastic content, if he believes in a set of sentences A, he also believes in any sentence p that is the logical consequence of A, i.e., he believes in any p insofar that $A \vdash p$, or equally, insofar that $p \in \mathrm{Cn}(A)$.

The most significant model in the theory of belief dynamics, namely the AGM model, is developed for operations on belief sets. Using this model, if we want to add a belief p to a given belief set K without removing any belief in K (while preserving logical closure), we can thus write the operation as $K + p = \mathrm{Cn}(K \cup \{p\})$, where '+' denotes the operation of *simple belief expansion*.[6] Using the set-theoretical tools, one can model more complex operations of belief revision. I will come back to this later.

---

[5] Alchourrón, Gärdenfors, and Makinson (1985).

[6] We use the *union* of two sets to describe the expansion of propositions. But it can also be modelled to coincide with possible worlds, in which the *intersection* of sets is used since the set of possibilities is diminished: For a belief set K and a sentence p, the simple belief expansion of

However, there are serious limitations on the applicability of the model of belief set: It sets a too high standard for commitment since it requires the agent to be doxastically committed to every sentence that is the logical consequence of the belief that the agent thinks he has. For example, if I believe that today is Tuesday, upon being asked if I believe that either today is Tuesday or Aristotle died in 322 BC, as a rational agent I should say yes, since this latter sentence is also in my belief set; but for the same reason, I should also say yes if being asked whether I believe that either today is Tuesday or Aristotle did not die in 322 BC. In fact, models for belief sets typically operate on logically infinite set of belief sentences or possible worlds. But since as finite intellects our discursive power is limited, a model of logically omniscient agent is hardly satisfactory and obviously does not go along with real situations.

Therefore, it is desirable for a theory of belief dynamics to distinguish the beliefs that are 'basic' to our doxastic commitment from those that are merely derivative of the basic ones. We can call the set of basic beliefs of an agent her *belief base*, which does not need to be deductively closed. If a set of sentences A is a belief base, for any sentence p we can make the distinction as follows:

p is a belief iff $p \in \mathrm{Cn}(A)$

p is a basic belief iff $p \in A$

p is a merely derived belief iff $p \in \mathrm{Cn}(A) \backslash A$

Belief bases has the advantage over belief set to fit with the phenomena of our belief dynamics. As a very simple example, consider two belief bases: $\{p, q\}$ and $\{p, p \leftrightarrow q\}$, they have the same logical closure and hence belong to the same belief set, nevertheless we consider them as indicating two different kinds of doxastic commitments—If the removal of belief p is needed, an agent having the latter kind of belief base must also remove q, but an agent with the former kind of belief base needs not to do so—Therefore, we can say that even though the two belief bases are statically equivalent, they are not dynamically equivalent.[7]

We can model the operations on given belief bases in a similar vein as the model for belief sets. But the difficulty is to answer the question: Which belief is a basic one? How to demarcate a set of sentences as a belief base from other derivative beliefs?

K by p will be a non-empty intersection of the two sets of possible worlds, i.e., $[K] \cap [\mathrm{Cn}(\{p\})]$. See Hansson (1999, 220f.).

[7] See Hansson (1999, 20).

In formal theories of belief dynamics, it is usually suggested that we can regard belief bases and their operations as implicit, hidden structures of one's doxastic commitment—in contrast to the explicit, overt structures of one's doxastic commitment described by belief sets. (We might compare this difficulty to our previous observation that the standard for determining the boundary of one's doxastic commitment is pluralistic and relative to the agent's discursive power.)

If we want to demarcate an agent's basic beliefs, we obviously need meta-level policy to operate on the agent's beliefs and their revisionary policies. Recently, Sven O. Hansson has proposed a metalinguistic belief predicate $\mathfrak{B}$ to describe the dynamics of belief revision.[8] The distinction between linguistic and metalinguistic doxastic operators correctly pinpoints the problem that the question of doxastic commitment is not primarily a question of description of the 'static' possession of sentences of a certain coherent set of sentences; rather, it is primarily a question of 'dynamically' adopting or undertaking a set of sentences that are the appropriate outcome of a revisionary act.[9]

Another issue about the theory of belief dynamics that needs to be noted is the relation between *belief revision* and *belief contraction*. For the AGM model and hence the mainstream of the theory of belief dynamics, the operation of belief contraction, which removes a sentence from the belief set, is regarded to be more primitive than the operation of belief revision. It means that the logician's task primarily concerns the modelling of belief contraction, while belief revision can be seen as a product of combining the operation of belief contraction with belief expansion—Revising a belief set K to include p consists in first contracting K by $\neg p$ and then expanding the set by p. If we use '$-$' to denote the operation of contraction and '$*$' the operation of revision, then we have a so-called 'Levi identity': $K * p = (K - \neg p) + p$.

Such an approach treating revision to be derivative has disadvantages: First, since belief revision is derivative, the operation of belief expansion does not have the

---

[8] Hansson (2014, 2016). For a sentence p in the object language in which beliefs are expressed, a truth-functional operation $\mathfrak{B}p$, which Hansson calls a 'belief descriptor', represents a metalinguistic set of sentences that are appropriate to be believed (formally: the metalinguistic set specifies the success condition of belief revision).

[9] Nevertheless, since this dissertation does not have the intent to handle axiomatic formalizations, for us introducing a metalinguistic belief predicate selecting potentially desirable outcome set that is stable and coherent out of a set of belief sets does not really alleviate the problem.

function of consistency-check. In other words, if we apply belief expansion separately, it might result in having contradicting beliefs in a belief set and thus fail to preserve consistency.[10] But then it is unclear how belief revision can be derivative of such a problematic operation. Second, the contraction in this primitive sense must also be 'pure' contraction, according to which we must remove a belief without believing in its negation. But it is hard to find an example for such pure contraction in real cases. In other words, in real cases the 'inclusion postulate' ($K - p \subseteq K$) for belief contraction can hardly be satisfied.[11]

Therefore, I hope it could now be convincing enough that instead of having belief revision being based upon two problematic operations, it is more advisable to regard belief revision as the primitive operation, while the two problematic operations can be seen as rather special, idealized cases of revision. In what follows, I will not discuss how to model the revision operator appropriately. I will only use the symbol for belief revision (*) for the sake of expediency of my own discursive description of belief integration as the consequence of reason-giving acts.

6.2.2 Operations of Belief Integration

According to the view of constitutive discursive agency, assertion is an act of discursive commitment to the appropriateness of the content at the context of assertion, whose evaluability condition is governed by the norm constitutive of the utterer's agency, which, in turn, involves the socio-historical determinations of the utterer's power. Therefore, the utterer's power is implied in the context of understanding and evaluating the assertion. Since assertion is primarily a self-evaluating act, the utterer's

---

[10] Accurately speaking, for belief set expansion, since we have the consequence operator Cn($K \cup \{p\}$), the outcome set will be logically closed. (If $\neg p \in K$ then $K + p = \emptyset$.) Things are more complicated in belief bases, where we will get different results if we do not accept the Levi identity but first expand and then contract the belief base instead. Cf. Hansson (1993).

[11] Related to this problem is the even more debated 'recovery postulate': $K = (K - p) + p$. Cf. Levi (1991, 131f.). Since a 'pure contraction' does not yield any new belief but only avoid falsehoods, it is not a knowledge-expanding operation. By contrast, the observation that belief revision is more fundamental than 'pure contraction' (which may be seen as rather a special case in a limiting situation) coincides, historically, with Hegel's idea of determinate negation, i.e. the idea that negations are inherently determinations, which, therefore, are nevertheless expansions of our knowledge, rather than purely-contracting and hence non-knowledge-expanding operations.

consciousness of her power is also part of her background beliefs implied in the context at which she asserts. Let c_i be the context of the assertion, and Γ_i be the set of the agent's background beliefs, i.e. the complex body of premises and constitutive norms implicit in the assertion. (Γ_i, therefore, can be seen as a belief base.) According to the dynamic view, asserting p at c_i is hence committing oneself to integrating p with what one is doxastically committed to, i.e., Γ_i.

Our task, now, is to account for the detail of the dynamics of belief-integration, which inherently involves change of context and hence one's background beliefs—When a reason P is given to agent S, S either will or will not adopt P as a belief she is committed to. In what follows I shall consider the belief integration as the consequence of the unqualified, the disqualified, and the qualified cases of reason-giving, respectively.

Let us start with the unqualified case first: In the unqualified case, upon a given reason P, S can accept P and agrees to assert it in p without experiencing any dissonance. Now suppose Γ_1 is the set of S's background beliefs before integrating P, while Γ_2 is the set after integrating P. S hence asserts p at c_2 with Γ_2 (i.e., $p \in \Gamma_2$). Since P induces no incompatibility, S can add P to her background beliefs without the need to change any of her preexisting beliefs. It thus exemplifies an operation of simple belief expansion:

(1) $\Gamma_2 = \Gamma_1 + p$

This is the case typically when p is deducible from Γ_1 (i.e., when $p \in \mathrm{Cn}(\Gamma_1)$). But as discussed earlier, a simple belief expansion can also take place when we add a piece of new information to our background beliefs—Like the example in §6.1.1, when John asks Mary to check the temperature in Boston for him, Mary tells him that it is 20. Due to Mary's trustworthiness implicit in the conversation, John can add the belief that the temperature in Boston now is 20 degrees Celsius also in a way of simple expansion, i.e., with minimum friction.

Nevertheless, in the last subsection I showed that the operation of simple belief expansion cannot preserve consistency on its own. In other words, it is not an independent operation; rather, we should regard it as only a special and idealized case of the more primitive operation of belief revision—i.e., as the revision with minimum to zero belief-contravention:

(1*) $\Gamma_2 = \Gamma_1 * p \simeq \Gamma_1 + p$

(1*) indicates that in non-deductive cases, the equation of revising Γ_1 by p is only *approximately* non-belief-contravening.

We can explain this proximity using a finer structure of being-told: If Mary tells John: "The temperature is 20", John's belief about what he has been told is p = 'The temperature is 20', while he draws the conclusion that q = 'The temperature *in Boston now* is 20 degrees *Celsius*' only by inference. Hence, we have something like:

(2) $q \in \Gamma_2$, while $\Gamma_2 = \Gamma_1 * p$

Strictly speaking, the part of belief integration that is unqualified is rather the inference to q, while the information intake itself represented by $\Gamma_1 * p$ always has frictions—Being-told concerns a rather asymmetrical structure of knowledge reception, which needs to be accounted for by the mechanism of qualified reason-giving (which will be discussed later).

This squares with the idea that the inferences about our empirical knowledge are always defeasible—An agent accepts a reason without qualification only provisionally so, and the acceptance can be overthrown—If Mary continues her sentence: "The temperature is 20 …… Fahrenheit.", John can no longer infer that q. Therefore, we say that John's inference from p to q is defeasible and nonmonotonic.

Now, let us turn to the belief integration in the case of disqualified reason-giving: In the disqualified case, when a reason P is given to S, S fails to incorporate P into her background beliefs set Γ_1, we may use a mark '*?–*' to indicate this kind of failure or confusion:[12]

(3) Γ_1 *?–* P

We may assume that all our courses of recognition start with more or less unexpectedness, i.e., encounter with disqualifications. But the consequence of such courses of recognition varies—Basically, three kinds of things might happen: One possibility is that the agent eventually succeeds to integrate P (i.e. carries out revision of Γ_1 by p) after some effort, and P hence ceases to be meaningless.

Another possibility is that the agent remains confused, i.e. fails to integrate P into her background beliefs—even after some effort; during this period of effort, if the

[12] Note that (3) is not a well-formed formula but only marks in our context the failure of constituting a conceptual content. Although I use the question mark here, it also cannot be equated with interrogation. (On interrogation cf. Appendix 3).

agent has had a change of background beliefs from Γ_1 to Γ_2, we can say that it remains the case even for Γ_2 that P is meaningless, thus: Γ_2 ?– P.[13] This hence shows a failure of discourse.

The third possibility is somewhat peculiar—As discussed, there is possibility that we understand disqualification in terms of context-shift—In such case, although S fails to make sense of P according to Γ_1 as (3) indicates, there could be a subordinate context d, at which P is sensical and it is discursively appropriate to assert p. Let Δ be the set of background beliefs implicit in d, Δ is hence a subset of Γ_2 implicit in a global context c_2, at which the agent succeeds to incorporate d as a shifted context:

(4) $p \in \Delta \subset \Gamma_2$

Therefore, we can regard context-shift as nevertheless a case of global belief revision and hence qualified as reason-giving. Its difference from the normal belief revision is that the integrated sentence p is always dependent on the subset Δ; namely, the sentence behaves like part of a 'belief package'.[14]

As discussed in §5.5, context-shift indicates faultless disagreements and mishaps. Now we can see how it works: Suppose that p is a line in Hamlet, Δ entails the belief that there is a performance of Shakespeare's play taking place. According to Γ_1 the agent does not know Δ and is hence confused by the encounter with sentence p; but with Δ incorporated in Γ_2 the agent thereby is able to make sense of p. Another example: Let p be the statement that the Americas is the Indies. An agent with Γ_1 knows of modern geography and thus considers p to be simply false. Now suppose that Δ entails beliefs about the history of human exploration of the world. If we incorporate Δ with Γ_1 in Γ_2 the agent not only knows of modern geography but also knows of the history of our geographical discovery, so that p put into context (e.g. the context at the time of Columbus) will be not simply false but perhaps conditionally appropriate (i.e. relative to historical context); and in this sense, p will be understood not as a genuine mistake but rather as a mishap.

---

[13] This is thus not restricted to the case that the agent rejects p by stating its negation according to Γ_2. It could also be that P is senseless or unintelligible to the agent.

[14] Technically, we might suppose that there is an operator '$\circ$' for revision by a set of beliefs—e.g. as meta-level belief revision—then we can understand context-shift as revising Γ_1 by Δ so that $\Gamma_2 = \Gamma_1 \circ \Delta$.

So far, I have discussed the belief integration as consequences of unqualified and disqualified cases of reason-giving. It has become clear that the understanding of any of these cases requires the understanding of the operation of belief revision by virtue of the qualified case of reason-giving. The mechanism of revising the background beliefs by integrating a piece of new information (e.g. $\Gamma_1 * p$) is presupposed by the unqualified and disqualified cases. However, it remains unclear how this mechanism could be really understood. Let us take a closer look at it in the next section.

6.2.3 Belief Revision and Counterfactuals

The qualified cases of reason-giving are discourses that are generally pedagogical or polemical. Such discourses induce genuine disagreements and demonstrations of mistakes. Therefore, the operation of belief revision as consequence of such kind of discourse involves both adding a new belief and abandoning old ones by admitting fault. It means that in such case there is certain sentence p that belongs to Γ_1 but does not belong to Γ_2. However, if we hold the relevant sets of background beliefs Γ_i fixed, it is unclear how falsehood or mistake could be understood since p is *always correctly* believed or disbelieved according to one's background beliefs *relative to* their very statuses of belief or disbelief concerning p—As we have seen, the shifts of contexts, faultless disagreements, and recognition of mishaps exactly exemplify such kind of relativity of contexts and their respective evaluative norms. But it does not apply to the cases of genuine disagreements. In other words, we cannot confuse the shifty relation between Γ_1 and Δ on the one hand with the non-shifty, continuous relation between Γ_1 and Γ_2 on the other.

We have discussed the genuine and the faultless disagreement in the last subsection. The moral of the discussion, I think, consists of two points: *First*, the intelligibility of mere mishaps is secondary to the intelligibility of mistakes—Recognizing a faultless disagreement is doing more, not less, than recognizing a genuine disagreement—For example, In the last subsection I suggested that when facing the statement "The Americas is the Indies", if we take its relative historical context into consideration, there is possibility that we recognize the statement as appropriate relative to its own historical context and only be viewed as a mishap, i.e. a merely indifferently different statement from our own statements about geography. However, a mishap as such is only intelligible when it is relative to a subordinate context that sanctions a shift

in epistemic standards, while the sanction of the shift is answerable to the norms and standards in the global context and hence is eventually liable to mistakes. *Second*, I argued that we can only understand genuine disagreement in terms of the agent's self-disagreement and hence in a sense in terms of regret. The speaker of a qualified reason-giving act hence intends the hearer to internalize the relevant norm thus given and to be regretful.

As noted, the regretfulness here does not refer to a sheer emotion but rather a feature of pragmatics constitutive of understanding genuine disagreement. But how can we understand it? One feature of regret is that it involves content that induces *contrast* in a discourse, i.e. a surprise to the agent's expectation. However, it does not help distinguishing between genuine and faultless disagreements—In both the qualified and the disqualified case of reason-giving, the hearer experiences some dissonance, which eventually turns into either genuine disagreement or a faultless disagreement.

There is yet another important feature when the agent is regretful about his previous belief(s), namely, the agent expresses the regret in counterfactuals—If the sun were not to rise tomorrow, we would be surprised. Our best scientists would start immediately to try to figure out what had gone wrong in our existing physical theories and beliefs that could not predict such an event happening. Conceivably, they might find out in the end that there was indeed a mistake or flaw in their theories which they should have avoided—They could have avoided it, but they regret that they did not. In such cases, the avoidance of mistakes is always expressed counterfactually.[15] Therefore, the recognition of mistake concerns not only description of the occurrence of

[15] What if there is no regret in failing to predict the disappearance of the sun? Imagine that eventually the scientists failed to comprehend this baffling phenomenon, and some alien creatures descended to the Earth telling us that all the astronomical phenomena we used to observe and theorize were in fact just their trick on us, hence that there never had really been a sun. What would we do then? We would pack all our previous beliefs about astronomy (and probably about physics in general) into a giant subordinate context and regard the history of our exploration of the universe in this context as not a mistake but a sheer (and sadly, pathological) misunderstanding. We would not express such misunderstanding as avoidable and hence no counterfactual statement would be necessary. Obviously, this is an extreme example. It seems simply unlikely that if the sun suddenly disappears we would just regard it as a mishap, for our knowledge and education aims exactly at preventing this kind of scenario from happening. At the global level, therefore, this is most likely to happen only in situations with severe anachronistic relocation.

dissonance or surprise in the agent's causal history but also the counterfactual, normative statement that the agent actually should not have surprised.[16]

Not all counterfactual statements can be used to express regret, though. It is helpful here to distinguish between two kinds of counterfactuals: the *backtracking counterfactual* on the one hand, and the *non-backtracking* or *simple counterfactual* on the other. And we express regret exclusively in terms of backtracking counterfactuals. We can distinguish the two through the following example:

Suppose someone is standing at the edge of a roof of a tall building as if about to jump off. The scene makes people crowd around on the street. Suppose you and I are also in the crowd. Now you say: "If he had jumped, he would have died."—This is a simple counterfactual statement. It says that given the counterfactual condition that the man had jumped he must be ignorant of the fatality of such a practice (so long as the man is not committing suicide). But consider that I might also reply you by saying: "If he had jumped, he would have lived."—The reason for me to say so is that since the man seems sufficiently rational and not suicidal, given the counterfactual condition that he had jumped it must be constitutive of his practice that he in fact knows something we don't, e.g., a life net or an inflatable cushion at ground level—This, then, is a statement of backtracking counterfactual.[17]

It is an oversimplification to comprehend backtracking counterfactual as the counterfactual that makes statements about the past. For example, suppose I used to have two hands but later lost my right hand, then I say: "Had I have two hands, I would be able to play this piece on the piano."—This is *not* a backtracking counterfactual, but only a simple counterfactual—Admittedly, from the context of my utterance, we know that I used to have two hands; but this only adds a bit anachronism and bitterness about my loss—Though we may say that in such case I express some 'regret' in a broad sense, this kind of 'regret' lacks the intrinsic normative character as the regret in admitting fault has—Equally, if the lottery I bought has rolled out its numbers 'UVW', and I then claim: "Had I bought the lottery ticket with numbers 'UVW', I would have won.", this

[16] I am largely indebted to Pirmin Stekeler-Weithofer's pragmatist interpretation of abductive reasoning. Cf. e.g. §7.5 of his manuscript, where he points out that the German word for necessity (*Notwenigkeit*) also expresses the urgent need to turn (*Not-wendig*) to the new solution.

[17] See Justin Khoo (2017, 841f.).

is also a simple counterfactual statement. It rather expresses a mere wish, which has no internal normative connection with my past. In fact, we can imagine that I might claim instead that I would have lost if I had bought the lottery ticket with numbers 'XYZ' (for any numbers that are not the winning ones). It just makes no difference here.

Therefore, the backtracking not only expresses counterfactual conditions that are about the agent's past, but more importantly, it also expresses them in an inherently normative way: While expression of a simple counterfactual exemplifies norms that are not constitutive of one's practice ('He would have died if he had jumped.'), expression of a backtracking counterfactual exemplifies norms that are thought to be constitutive of one's practice ('He would have lived if he had jumped.'). Later, I will further show that the backtracking counterfactual condition is inherently power-related.

6.2.4 Historical Revisionism and the Movement of Content

We set our task to account for the mechanism of belief revision. The discussion in the previous subsection shows us that counterfactuals could be useful for this task. Moreover, it shows that not any counterfactual will do the work but only those that are inherently normative, i.e., those that invoke constitutive norms of the agent's practice. We called such kind of counterfactuals the 'backtracking'.

It is important to incorporate the aspect of backtracking into the counterfactual analysis of agency. Nozick once adopted a counterfactual analysis in his account of knowledge, which he summarized as follows:[18]

> *(a)* S believes, via method or way of coming to believe M, that *p*.
> *(b)* If *p* weren't true and S were to use M to arrive at a belief whether (or not) *p*, then S wouldn't believe, via M, that *p*.
> *(c)* If *p* were true and S were to use M to arrive at a belief whether (or not) *p*, then S would believe, via M, that *p*.

The method M is in a sense comparable to the agent's background beliefs and norms Γ in our context. Nozick argued that *(a)*, *(b)*, and *(c)* are conditions constitutive of one's knowledge. I cannot discuss his argument in detail here, but we can see that without a proper differentiation of the backtracking counterfactuals from the simple, non-backtracking ones, Nozick's account lacks the resource to resist deflationism, i.e., the

[18] Nozick (1981, 179f.)

mutual definition between the truth condition of one's belief on the one hand and the application condition of one's method M to arrive at that belief on the other.[19]

As I have discussed in Chapter 5, there is a danger in considering the evaluability condition (either of the appropriateness of a content or of the appropriateness of a performance) of an assertion as being glued at a given context, which makes the comparison and evaluation of assertions across contexts either trivial or impossible. Besides, I have been proposing a dynamic view of discursive agency that I believe can overcome this difficulty. For a dynamic view, one's transition of her background beliefs form Γ_1 to Γ_2 cannot be reduced to two sets of sentences that are intrinsically irrelevant to each other, nor be reduced to the logical 'annihilation' of Γ_1. Rather, as I suggested in §5.5–5.6, we should comprehend the transition as a movement, which consists of two mutually irreducible orders: the *order of recognition* and the *order of explanation*. The order of recognition is from context-preservation to context-revision, while the order of explanation is from context-revision to context-preservation. Now we can have a more detailed presentation of this movement: Let Γ_1 be the set of background beliefs and norms implicit in the context c_1 before revision, and Γ_2 be the set of background beliefs and norms implicit in the context c_2 after revising Γ_1 by p. (Hence, Γ_1 indicates a low evaluative standard, while Γ_2 indicates a high evaluative standard.) The two orders of the movement can be depicted as follows:[20]

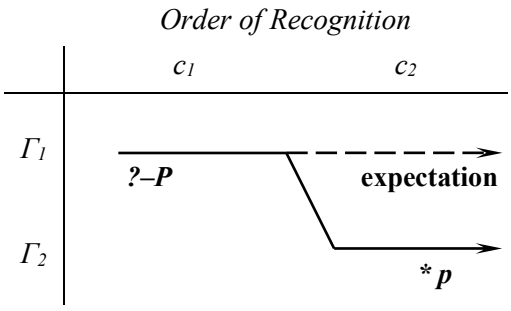

Fig. 6.1

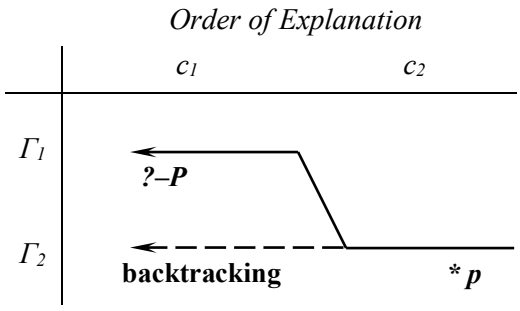

Fig. 6.2

Fig. 6.1 shows that in the order of recognition, the agent has gone through a surprise, i.e. dissonance in expectation: Upon encountering P he is confused at c_1 and cannot

[19] A detailed discussion on this account cf. Kripke (2011).

[20] Note that the diagrams below do not represent any metaphysical model of the branching of worlds, but only indicate the different normative statuses from aspects of the movement of belief revision. For an Anscombean approach to temporal unity cf. Lavin (2015).

make sense of it. But the agent's original expectation according to Γ_1 is abandoned after successful revision by p at c_2. Hence the agent can speak of the falsehood of the Γ_1-expectation at c_2. (Suppose that Γ_1 entails sentence q that is incompatible with p, the agent can then speak of q in a counterfactual manner, like that it would have been inappropriate if he had asserted q at c_2—Note that this is just a simple counterfactual and is not inherently *normatively* connected to his past background beliefs Γ_1, since we can imagine that the agent can speak of the inappropriateness of any set of beliefs Γ_i that entails q at c_2.) The order of recognition helps the agent to foreground Γ_1 (rather than any other Γ_i) because this is his recognitive past.[21] But to really establish the normative connection between the agent's past and current beliefs, we need to have the order of explanation in view, hence—

Fig. 6.2 shows the agent's retrospective explanation of his pre-revision normative status. He brings up his previous failure to integrate p into his background beliefs Γ_1 and admits fault. But as discussed, this is only sensible when the agent also expresses regret—in terms of the backtracking counterfactual. In backtracking, the agent claims something like that it would have been appropriate if he had asserted p at c_1 according to Γ_2—It is not a simple counterfactual since the agent also claims that the epistemic standards and norms for asserting p is supposed to be constitutive of his agency at c_1.

To sum up, we can see that confusion, false expectation, successful revision, and backtracking constitute four essential and irreducible normative statuses in the movement of belief revision. The movement is manifested in two orders: the order of recognition and the order of explanation. Moreover, they are two aspects of the same, unitary movement—while the order of recognition foregrounds the movement's aspect of causality, the order of explanation foregrounds the aspect of normativity—and we cannot understand either of the two aspects without understanding the other. On the one hand, an agent's explanation of regret loses focus if without the help of the aspect of his recognitive transition to foreground his past; On the other hand, with the help of retrospective explanation the agent's recognitive transition has not only a past and a

[21] Yet it does not mean that the recognitive past and its original expectation itself is not normative. If q is the expectation according to Γ_1, it itself is the conclusion of an inference to the best explanation before revising by p, the assertion of which implies norms constitutive of the agent at c_1.

present, but a history—The agent's regret or backtracking of past cognitive failures is not expression of a sheer emotion of wish that one would have done otherwise, but essentially a way of looking into our history.

The dynamics of the movement of content or thought illustrated above provides a conceptually basic unity that can be analyzed in terms of the differentiation of contexts and their corresponding constitutive norms. However, I would like to emphasize that the illustration above by no means intends—*pace* my use of the symbol for belief revision (*) in discussion—to give or to define a *function* of belief-revision (typically parametrized by contexts as arguments, which, as argued, is rather a problematic approach to understanding context-sensitivity).[22] The movement is conceptually basic because the differentiation of contexts and norms qua internal differentiation (i.e. qua self-disagreement) requires understanding of the unity of the movement and its two orders. Note, however, that although it is plausible for us to claim that the orders of the movement correspond to orders in time and hence suggest that the movement is temporal, it does not presuppose any metaphysical framework of temporality. Of course, more work can be done on topics of temporality and historicity, etc., with regard to the dynamics of belief revision. But it exceeds the task set for this dissertation.

The most important consequence of such a dynamic view is that preserving a context and revising it are not two steps, neither are they two separate operations. The historical representation (or to use Brandom's term: recollection[23]) of our previous knowledge and doxastic contents is inherently a way for us to explain and evaluate them in a distanced manner as acts of discursive agency with qualification according to our current norms and standards[24]—Let us call it a *historical revisionist* view.

The historical revisionist view should be distinguished from a 'sheer revisionism'—or, as what I would like to call: a *historical nihilist* view. The difference between the historical revisionism and the sheer, historically nihilist revisionism is that

---

[22] Neither am I here able to further investigate how to provide a more systematic model of the dynamics of belief revision as depicted above but only give a discursive description of the unity of the movement of revision. In the next section, I will provide further argument that the manifestation of our discursive power should not be grasped in functional terms.

[23] Cf. Brandom (2015) chap. 7.

[24] And again, a sheer preservation of context and content, i.e. a context-shift, requires us to do more rather than less than context-revision.

for the sheer revisionism the evaluation of a distanced assertion is carried out irrespective of the assertion's underlying epistemic standards and norms in its context of use. Quite the contrary, according to the historical revisionism, revision cannot be done without basing on the continuity with previous beliefs and their constitutive norms—For the explanation and evaluation of a distanced assertion we must take the agent's socio-historically mediated discursive power into consideration.

For historical revisionism, in other words, it is constitutive of the act of belief revision that we must distinguish a backtracking counterfactual from other counterfactuals, i.e., distinguish a (regretful) mistake from other errors and mishaps. As discussed in Chapter 5, deflationists (and social constructivists such as MacFarlane) tend to account for assertion-evaluation in terms that we communicate in a most efficient way that requires minimal contextualization and demand of intellectual resource. However, they adopt such an approach at the cost of failing to account for the conceptual significance of discursive commitment, accountability, mistake, and regret. We may regard their approach as an example of historical nihilism—For historical nihilists, if commitment, accountability, mistake, and regret could be constitutive of assertion-evaluation at all, they do not play any conceptually significant role. Hence they do not have the resource to distinguish between admitting a mistake and rejecting a proposition.[25]

Last but not least, we can see from this historical revisionist perspective that the movement of belief revision is inherently a pedagogical movement. An agent's history of discursive practice can be seen as a process of *self-education*—In unsuccessful cases, the relevant genuine disagreements could not be resolved, and the discourse remains being polemical. In successful cases, the disagreements are resolved; more epistemic standards and norms are internalized and become constitutive of one's discursive acts. In such cases, the agent's knowledge grows.

[25] For example, by admitting fault, one is regretful of his mistaken assertion; but since he is making a discursively appropriate assertion about the falsehood of an assertion, as a historical nihilist the agent could simply bootstrap himself out of regret—Therefore, for the historical nihilists the accountability and regret are just conceptually trivial. A similar and more detailed discussion on this topic cf. Moran (2001, 174ff.). For historical revisionists, by contrast, a mistake could be forgiven, but it is not identical to being forgotten or ignored.

6.3 The Discursive Power in Assertion

In this chapter, I delineated the dynamics of belief revision. It shows that our practice of thinking essentially consists in the discursive movement of giving and asking for reasons (or equally: movement of *reasoning*), which concerns changes in constitutive norms—An unqualified discursive act can be understood as part of this movement, just like that an assertion can be understood as part of an inference or reasoning—It is beyond the scope of this dissertation to dig deeper into the issues of inference. However, it suffices here to show that the solutions to the central problems of evaluability of contents and of discursive agency reside in regarding inferences as movements of contents that are irreducible and conceptually basic.

We understand the qualifications of our agency in terms of the acknowledgement of mistake (which, clearly, is itself not a mistake), which is performed under the form of the practice of thinking as an act of discursive commitment. On the other hand, however, we do not understand the performative act of discursive commitment, i.e. the full-fledged assertion, if we do not understand it as belonging to the discursive practice of reason-giving, i.e. as a self-educating movement, which, in turn, comprises epistemic frictions and socio-historical qualifications of our agency. In sum, we can make three observations of this dynamic movement: *a)* Metalinguistic modeling does not seem to help solve the problem; *b)* The agent of the practice of thinking expressed in a content in the first person rather refers to one's background knowledge, which is subject to 'historical revisions' in a peculiar self-constituting way that concerns the phenomenon of normative backtracking; and *c)* The normative backtracking is irreducible because it is more basic than other context-relative interpretations of content.

We generally see a movement or a motion as the exercise or performance of a capacity or a power. Since I suggested that an agent's act of assertion can be seen as performance of her *discursive power*, in this final section I would like to say a bit more about the idea of discursive power.

When I introduced the notion of discursive power in §4.2.4, I distinguished it from the notion of a *regular capacity*. Unlike a regular capacity, our discursive power is autonomous and spontaneous—It is not a sheer potentiality, and hence does not signify any specific function—A regular capacity (or potentiality, or disposition), on the contrary, can be characterized in terms of a function: For a capacity ξ, we may say that

it has a functional manifestation M, so that $\xi(x)$ = M. For example, for something *x* to have the capacity of being soluble in water is for it to have the manifestation of dissolving in water—However, for it to perform the capacity and hence to achieve such manifestation it must satisfy certain conditions, e.g. that it is placed in water at an appropriate ambient pressure and temperature, etc. These are conditions for the instantiation of the capacity's function of manifestation. It seems more appropriate, therefore, to formulate the manifestation conditions somehow as follows:

> (5) For all *x*, time *t*, and circumstance *c*, if *x* is placed in water at *t* under *c,* then *x* is soluble in water at *t* under *c* if and only if *x* dissolves at *t* under *c.*

The formulation of (5), however, indicates that the manifestation of *x*'s capacity of solubility is conditioned by factors that are not constitutive of the capacity; namely, to instantiate a manifestation of such a capacity it depends on an external condition to *activate* it (e.g. to place *x* in water).[26]

We have discussed the problem of instantiation with respect to the issues of motivating force in §5.2, which is clearly related to the problem of capacity-activation: Any capacity whose performance is not spontaneous but needs external trigger or motivating force has the problem (e.g. the capacity for a car to move, or the capacity for a computer program to win a game of chess, etc.)—As shown in §5.2, we can imagine a computer program equipped with all the legal moves of chess and their consequences; but it nevertheless lacks the disposition to carry the game towards a winning position (rather than a losing one). Therefore, though we can say that the program has a capacity to win a game of chess, and that it can perform such capacity in a movement (a series of moves) towards winning, it lies outside of its capacity to be activated or motivated to win—Although we can of course further parameterize the activating condition of that program and devise a function favoring the winning position over the losing ones, the conception of such a meta-capacity of activating a capacity's manifestation is obviously regressive (i.e. we need another activating condition to activate the manifestation of this meta-capacity, etc.).

---

[26] This is Carnap's classic definition of disposition or power in terms of conditional sentences. A more detailed discussion of its problem cf. E. J. Lowe (2011). In fact, we can see that Nozick's counterfactual analysis of knowledge resembles the counterfactual variant of Carnapian conditionals: e.g. If *x* were placed in water at *t* under *c*, then *x* would dissolve at *t* under *c*.

Activation is needed for manifesting a capacity because a capacity may be either inactive or active. Yet even if a capacity is activated, it may still either succeed or fail to achieve its manifestation. Therefore, the instantiation of a capacity's manifestation is not confined to the question of activation—It also involves the question how to deem a manifestation as *successful*.

Living creatures can perform many of their capacities spontaneously, so that they can activate themselves to instantiate the manifestations of those capacities—but they might nevertheless fail to achieve manifestation. For example, a moth has the capacity of lunar navigation, and it can spontaneously activate itself to manifest the capacity. However, it could be that the poor creature happens to fly into a flame or a blacklight and dies, hence fails to manifest its capacity of lunar navigation anyway. Here is another example: A ball-juggler has the capacity to juggle four balls, and he might well be motivated to perform such a capacity. But imagine there is a sudden gust of wind that blows a ball away, so that he fails to manifest his capacity regardless that he has activated himself to do so. These examples show the problem of capacity-manifestation beside activation: If we regard the manifestation of a functional capacity as implying a set of constitutive norms, then this set of norms is indefeasible. Thus, once a capacity is activated in an instance of performance, the evaluability condition of that instance is also settled. The norms constitutive of that act, in other words, is conceptually 'exhausted' by the functional determinations of the capacity the instantiation of which that act is taken to be[27]—Though activated, the instance of performance is nevertheless inertia-like, and its success is based on contingent factors.

For a regular capacity, whether it is active or inactive and whether its manifestation is successful or unsuccessful are two separate questions. For example, although a sleeping or resting moth still has the capacity of lunar navigation, its capacity is inactive. But when it flies into flame, its capacity is activated but failed to really manifest. We may say that in the former case the moth has a *first-level capacity*: it is able to be able to navigate by moonlight; while in the latter case it has a *second-level capacity*, namely: it *is* able to navigate by moonlight—Such distinction is common in Aristotle's writings. (Moreover, it also corresponds to a significant distinction in Hegel's philosophy: We can say that at the first level the moth has the capacity in itself

[27] Cf. Zhan (forthcoming).

(*an sich*), while at the second level it has the capacity for itself (*für sich*))—However, even this second-level capacity is still different from our discursive power.

We can compare the distinction between discursive power and regular capacity to the Aristotelian distinction between *energeia* and *dynamis*.[28] The capacities of being soluble in water, of lunar navigation, and of ball-juggling are hence *dynameis* but not *energeia*. According to Aristotle's analysis, the manifestation of a capacity is a movement which, in fact, transcends that capacity—For *x* to manifest its capacity of solubility it needs to dissolve in water in a performance, so that *x* changes from being soluble to being dissolved; but being dissolved means that the capacity of solubility has accomplished its performance and thereby expired. (Similarly, in a performance of manifesting the capacity of ball-juggling the juggler changes from being able to juggle the balls to having juggled the balls.) Since the judgment whether *x* has dissolved cannot circularly depend on *x*'s solubility, the concept of being dissolved lies in this sense beyond the capacity of solubility—To use Aristotelian words, the *dynamis* for *x* is to become something other as such.[29]

The discursive power is not a *dynamis*. Unlike a *dynamis*, whose successful manifestation results in something other than itself, the performance of the discursive power as a movement has its result always lying within itself—As argued, it is neither a specific capacity to follow certain norms, nor an overarching meta-capacity of legislation that brings some norms into being as authoritative, whose manifestation does not concern the following of those norms. Generally, the discursive power does not serve as any specific function or follow any specific norm, and hence it would be problematic to say that it is a power *to φ*—It is not a power *to utter* this or that; rather, it is a power *in performing utterance*—Therefore, to activate such a power *is* to deem the manifestation of it as successful.

Admittedly, it might seem plausible to claim that, just like for a regular capacity, activation and manifestation-success are also two separate questions for the discursive power: We distinguished the performance of discursive power into the disqualified, the qualified, and the unqualified kinds, and this threefold distinction might seem to coincide with the two questions: I.e., there is a first-level distinction between being

[28] On Aristotle's distinction of the two cf. Jonathan Beere (2009), esp. cf. p.195 on the primacy of *energeia*.

[29] Cf. Kosman (2013, 49f.).

pathological (disqualified, inactive) and being non-pathological (active) on the one hand, and a second-level distinction between being successful (unqualified) in manifestation and unsuccessful (qualified) on the other hand. As shown below:

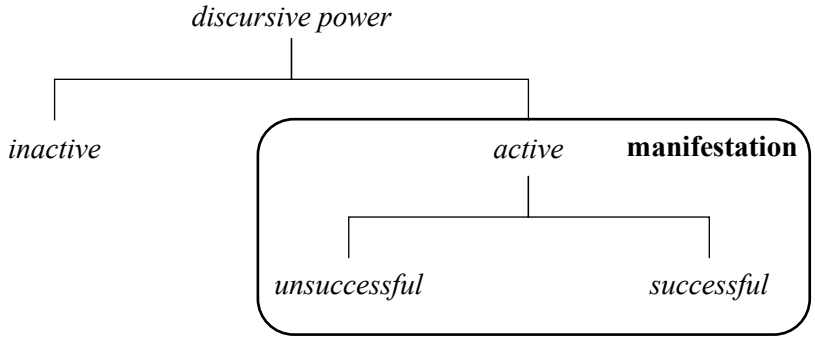

Fig. 6.3 (A Wrong Frame!)

However, as we have seen, if activation and manifestation-success are separate questions for the performance of discursive power, it would mean that the norms for use (i.e. for being discursively active) and norms for evaluation (i.e. for being discursively successful) could never be unified—It would mean that on the one hand we do not understand how the norms for evaluation can be applied in an instance of active performance of the discursive power, and on the other hand we do not understand how the norms for use can be applied with the consciousness of its appropriateness according to the norms for its discursive success. In fact, according to the dynamic framework introduced in this chapter, for the discursive power the distinction between successful and unsuccessful manifestation can only be understood in terms of a movement that is not different from the movement of activation—For an unsuccessful manifestation of discursive power, the power's pathological inactiveness must be in a sense recognized and in a sense overridden.[30]

It might seem ungrammatical or regressive to say that activation is a movement, i.e. to say that there is a movement from being inactive to being active, since a movement, by definition, is considered active. But it need not be regressive if we view the discursive power's movement as conceptually more basic and primitive than its being active or inactive. This may require us to reconceive the notion of movement and

[30] This is related to the debate over the so-called 'transformative theory' of human rationality. Cf. e.g. Boyle (2016). Since the transformative theory seems to claim that the manifestation of human rationality is logically exempt from pathological inactiveness, I am inclined to consider the transformative theory as problematic. Cf. Gobsch (2017).

of activeness in general. I cannot further pursue this topic here. But at least there is no serious obstacle, so far as I see, to conceive the performance of discursive power as a conceptually primitive and irreducible movement, i.e. a self-educating movement of building up the relevant agent's discursive power through resolution of disagreement—We do not need a meta-power to do so because the discursive power itself can be seen as self-activating and self-constituting.

The manifestation of an agent's discursive power is a movement that turns certain norms and epistemic standards into being constitutive of discourse; namely, it turns the norms into the agent's know-how. In a sense, we can say that a qualified performance of the discursive power is partially active and successful (according to a lower epistemic standard, say, Γ_1) and partially inactive and unsuccessful (according to a higher epistemic standard e.g. Γ_2). According to the dynamic framework, therefore, discursive activeness and discursive success are not really two distinct levels of manifestation of the power; rather, they are essentially the same movement (or maybe we want to call it: self-movement) with a multiplicity of degrees of activeness and success, which could perhaps be roughly illustrated in the following Figure 6.4 (the shaded arrow indicates the direction of the increasing level of epistemic standards and hence the growth of power).[31]

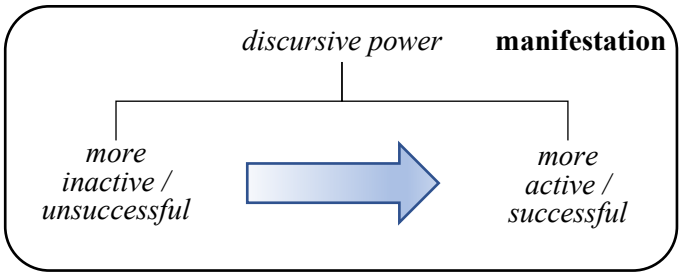

Fig. 6.4

[31] Discussion on this issue in the Aristotelian context cf. Kosman (2013), e.g. p. 60f.

Appendix

In this dissertation, I have made a broadly Austinian, act-theoretic approach to assertions and thoughts. According to the act-theoretic approach, the demarcation of thoughts consists in the demarcation of the pragmatic unity of speech acts; moreover, I argued that our speech acts are inherently illocutionary and committal, and our thoughts can be seen as inherently assertoric. This helps us account for a threefold distinction: *a)* Full-fledged Austinian assertion, namely illocution; *b)* Non-committal, non-context-shifting locutionary act; and *c)* Non-committal, context-shifting locutionary act. Both of the latter two kinds are unasserted, i.e. distanced, and hence explanatorily secondary, roughly corresponding to the Austinian rhetic act (typically the indirect speech) and the Austinian phatic act (typically the 'direct speech report'), respectively.

Admittedly, such threefold distinction—or more generally, the twofold distinction between committal and non-committal speech-acts—cannot cover all the fundamental philosophical features of our speech acts and our practice of thinking. Beside truth and meaning, our speech is rich in structure, syntax, and mode. They are connected with many other notions and issues that I have not dealt with in this dissertation: For example, I have not tackled the problems of temporality and modality. And though closely related, neither have I undertaken to propose a general theory of predication. It also relates to problems of understanding negation (negative predication like 'S is not P'). Furthermore, once we get down to the issue of predication, we need to account for the internal structure of a sentence, and then, we also need to explain how, say, complex and embedded sentences like '*p* or *q*' should be understood.

Questions like negation, embedment, and structured thoughts are difficult. Each of them demands detailed discussion and careful handling. This Appendix does not serve as an in-depth examination of any of these issues but only aims to show that their potential answers does not need to be at odds with my act-theoretic, commitment-based account of assertion—At least, I do not think an account of predication and structured thought will bring forth challenges to the act-theoretic approach any more than the challenges to those non-act-theoretic approaches like semantic minimalism. Therefore, I

think we can rather treat these questions as future topics that could be further investigated under the act-theoretic approach.

In what follows, I begin by briefly delineating the problem of negation. I will indicate that the problem of negation coincides with the problem of distanced thought (in the non-context-shifting sense). Then, I shall say something about reasoning, i.e. inference. Since our thoughts and assertions are inherently discursive, understanding reasoning, I maintain, is crucial for a thorough comprehension of the nature of assertion, and I think the dynamic treatment of discursive commitment and belief integration, which I have presented in this dissertation, coincides with the fundamental idea of reasoning.[1] At last, I shall briefly discuss the interrogative and the imperative expressions. These expressions are usually considered to be inherently non-assertoric—In other words, they appear to be lying outside of our committal–non-committal distinction—But I contend that they should not be conflicting with this distinction.

1. Negations

Generally, I think my account of non-committal, non-context-shifting contents is compatible with a proper treatment of negation. As argued, the non-committal, non-context-shifting content implies genuine disagreement. Since a genuine disagreement contains indication of error or mistake, it can be seen as an expression with negative content. To compare negation with genuine disagreement, however, it might be helpful to distinguish two kinds of negations: the 'predicative' kind and the 'sentential' kind.

By a 'predicative' negation I mean the negation that concerns explicit predication of a 'negative property', i.e. the negation that is 'sealed inside' the predicate—It can thus be written as 'S is non-P' instead of 'S is not P', for the use of a hyphen is deemed to indicate that it is rather the 'non-P' as a whole that is taken as a predicate—In this case, the whole sentence can nonetheless be seen as a fully committal assertion expressing affirmative predication. Since the negation only consists in the content of the predicate, it can be substituted with another positive predicate while its truth-condition remains unchanged. For example, if I say that the carpet is non-red, it

[1] Also, I contend that the act-theoretic approach to assertion can be made compatible with analyses of structured thought by adopting a proper treatment of reasoning.

implies that I hold a clear and definitive belief that the carpet has another color, say, purple. In this case, I can be seen as rather making an original assertion that the carpet is purple, which leads to the conclusion that it is non-red given the two colors' incompatibility with each other.

The 'sentential' negation, on the other hand, negates the whole sentence. Rather than predicating of a subject a negative property, it expresses failure of predication all together. Therefore, the 'sentential' negation is polemical in nature. In this sense, by saying that S is not P I am rather expressing that it is not the case that S is P (i.e. I am expressing *not p*). For example, if I say that the carpet is not red, I am simply denying that the terms 'carpet' and 'red' could be joined together to form a meaningful predicative sentence given the context, and my denial does not imply that I have any clear and definitive belief that the carpet has some other color.

Although I cannot give a comprehensive discussion of the issues at stake here, I believe the 'sentential' negation is primary for understanding negation—In short, while a 'sentential' negation expresses incompatibility between two sentences given the context, a 'predicative' negation expresses incompatibility between two predicates. However, just like that a 'literal' content cannot be semantically autonomous but its truth-condition is always dependent on its context of use, the incompatibility between two predicate terms, say, red and purple, is always context-dependent, too. Therefore, to understand negative property, I think we have to first understand how a sentence could be discursively inappropriate and genuinely mistaken due to its conflict and disagreement with another, not the other way around.[2]

Frege's idea of negation also refers fundamentally to the negation that takes place at the level of a whole proposition. To negate a proposition is to say that the proposition is false. Since it is the unity of the proposition that is negated, it seems to suggest that a false proposition is not a unitary proposition at all. But it then brings forth the challenge to explain how a proposition or content, with its unity being negated, can nevertheless be demarcated as judgable comparing to simply nonsensical occurrences—In fact, the intelligibility of falsehood is a question that can be traced back to Parmenides. But from an act-theoretic, 'top-down' pragmatic view, we can see that it is

[2] McDowell claimed in idem (1982), without much explanation, that expressions of falsehood work at a 'sub-semantical' level. I guess he also meant to point out the priority of the 'sentential' negations over the 'predicative' ones.

the same as the question of distanced analyzability, which we have treated in Chapter 6 under the dynamic, historical revisionist approach to qualified acts of reason giving and belief integration.[3]

2. Assertion and Reasoning

Assertion is a discursive act. And if we take discursivity to be equal to inferentiality, assertion is also an inferential act. This is a natural idea since an assertion can be seen as part of an inference—It is not accidental, after all, that Frege's assertion-sign has been used in logic as a symbol for consequence (the 'turnstile')—In other words, making an assertion can be seen as presenting the asserted content as conclusion of an act of inference or reasoning.

Apparently, though closely related, assertion and reasoning are two different topics—An act of reasoning concerns *multiple* assertions. A syllogism, for example, typically involves three assertions: the major premise, the minor premise, and the conclusion. But essentially, they are not two different kinds of acts. What holds the three assertions in a syllogism together is nothing but the assertoric force that indicates the agent's commitment to the truth of the conclusion of the syllogism.

Therefore, just like what Frege's Insight contends, i.e. that we must differentiate force from what is part of the content, there are reasons for us to equally differentiate the *logical form* that lies in the background of an inference and the expression of formality within an assertion. For example, the conditional sentence 'If P then Q' is expressed in a single assertion. In this manner, 'P' and 'Q' are parts embedded in a unitary content in form of the so-called 'hypothetical judgment', instead of an agglomerate of two separate propositions. (According to Frege's symbolism, they are attached to a singular judgment-stroke.) Thus, the embedded 'P' and 'Q' are different from the logical connection between the two sentences *p* and *q* in an inference.[4]

---

[3] Cf. also Hanks, who seems to be also equating 'sentential' negation with general cancellation. See idem (2015, 98–100).

[4] The same goes to sentences with other connectives such as conjunction and disjunction: "the conjunction is a single proposition, a logical unit, not a pair of separate propositions"; otherwise, "if you recognize conjunctive propositions as a kind of proposition, you may as well say, Mill remarked, that a team of horses is a kind of horse or a street a kind of house". (Geach 1965, 453).

From this point of view, we can see why the Frege Point is confusing—Its idea can be roughly put as follows: Consider the example of modus ponens reasoning: In order to recognize the validity of inferring from (i) If P then Q, and (ii) *p*, to (iii) *q*, one must see the same content *p* as occurring asserted in (ii) and unasserted in (i)—i.e. as expressing 'If *p* then *q*'. However, as indicated, 'P' embedded in a sentence is different from '*p*' as a complete assertion. That is why the premises of an inference are different assertions—The identification of 'P' with '*p*' does not occur automatically but only so when one draws the conclusion, i.e., only when one holds the two together in an act of discursive commitment. As I have discussed in Chapter 6, our inferential acts and acquisition of beliefs are generally nonmonotonic and defeasible. And hence we should not take the success of an inference for granted.

Admittedly, 'P' as part of a sentence is not 'simply' different from the whole sentence *p*. 'P' is not any part of a speech, say, the string of letters "resident" in the sentence "If Obama is the US president, he is a US citizen". Rather, to be identified with an assertion in a modus ponens reasoning, 'P' must be seen as an embedded sentence. In this regard, the Frege Point is nonetheless correct in observing that there must be a close relation between (*a*) the embedded speech in a compound, structured thought on the one hand, and (*b*) the indirect speech in a distanced, non-committal presentation of thought on the other. However, it is mistaken for the Frege Point to assimilate the two to an allegedly semantically self-contained content. The recognition of (*a*) as a content identifiable as (*b*) does not occur automatically but requires reasoning.

It is beyond the scope of this dissertation to provide a theory of compositionality that may help stipulate the syntactic and semantic rules for generating sentences like 'If P then Q' and explain how they govern our inferences, which may turn 'P' or 'Q' into assertions. But the basic idea, I believe, should be at least clear to the extent that the recognition of 'P' as an embedded sentence rather than a 'mere' part of a sentence is an act carried out in the conclusion of an inference. Since the conclusion of an inference is another assertion different from the premise, the analysis of intra-sentential composition does not supersede the discursive priority of sentences. Therefore, the problem of compositionality is not at odds with my act-theoretic approach to assertions.

It is one thing to argue what must be considered concerning reasoning in order for us to understand assertion, and another to argue what must be considered concerning assertion in order for us to understand reasoning. And since this dissertation focuses

only on assertion, it suffices for us to see that assertions are inherently inferential—As I have shown in Chapter 6, the dynamic picture of reason-giving and belief integration conforms to the description of defeasible reasoning.[5] In fact, in their basic pragmatic aspect—i.e. in the aspect of discursive commitment and of the revision of contexts and beliefs—the acts of assertion are not different from the acts of inference.

3. Interrogatives and Imperatives

In §4.1.3, I have clarified the relation between assertion and illocutionary points. I argued that assertion is illocutionarily basic—That is, according to my broad-brush Austinian view, assertion is not just a peculiar mode of presentation; rather, it *is* illocution and hence fundamentally differs from other modes of presentation. And since assertion is illocutionarily basic, we can simply focus on the assertoric–non-assertoric (i.e. the committal–non-committal) distinction without worrying about specific presentation modes.

But for some, there might be some modes of presentation that seem incompatible with my broad-brush, assertion-centered approach. Hanks, for example, despite his act-theoretic view of the explanatory priority of assertions over contents that largely agrees with my approach, maintains that questions and commands are nonetheless completely different kinds of speech acts that can be classified neither as acts of assertion nor as acts of distancing. However, while I agree that speech acts like commands and questions have complex pragmatic effects and cannot be *directly* classified into either assertion or distancing, I do not think we thereby need to conceive them as belonging to categories of their own kinds independent from the framework of assertion and discursive commitment. As Dummett once argued:

> In 'Über Sinn und Bedeutung' [Frege] explicitly repudiates the view the any sentences other than assertoric express thoughts; rather, he regards the difference between assertoric, interrogative, imperative and optative sentences as a difference in their *sense*

[5] The dynamic picture indicates a process of internalization of norms: In such a process, a norm may be in a sense non-constitutive and in a sense constitutive of the discourse—We can see that this also coincides with the fundamental feature of inference: the premises of an inference are in a sense individual assertions and hence individual acts independent from the conclusion, but in a sense constitutive of the conclusion in so far that the inference as a whole indicates a unitary discursive act.

> rather than in the force attaching to them. Thus he says that, just as assertoric sentences express thoughts, so interrogatives express questions, imperatives commands and optatives wishes. This view we may regard as definitely wrong […] (1981, 307)[6]

Commands, suggestions, questions, etc., each of these requires analysis of its own complex character in expression. Such character brings enrichment on top of the basic framework of commitment—They usually involve a complex blend of committal and non-committal statuses implied by the context. Take interrogatives for example: An interrogative could be analyzed either as a request to be told, or as expressing a desire to know (e.g. wondering), or as an expression of sheer confusion, or expressing a rhetorical question and hence implies disagreements, or a certain combination of any them, etc.[7] But these enrichments do not come into conflict with our basic act-theoretic framework of discursive commitment, disagreement, and pathology.

Take the imperatives as another example. I believe we should take the imperatives as inherently an act of reason-giving and hence as basically committal and assertoric (and fits the characterization of the act of reason-giving illustrated in Chapter 6)—The aim of a command is to give reason to the hearer—Rather than threatening or coercing the hearer, the speaker of an imperative sentence should intend the hearer to understand and accept the reason given by the sentence. In other words, the speaker should be guiding rather than 'goading' the hearer into accepting a reason.[8] Perhaps the imperatives are more involved with speech acts of practical nature rather than theoretical. But as I pointed out in §4.1.2, both theoretical and practical knowledge

---

[6] Cf. also Dummett (1993a, 152ff.).

[7] The U.S. Supreme Court once ruled that when a police officer says "May I look in your bag", that officer has issued a request—As Justice Scalia argued, the officer "has made it very clear that he's asking for your permission" (U.S. v. Drayton, 2002). But it is not self-evident whether a sentence in an apparent interrogative form expresses—as Justice Scalia maintained—request for permission. Actually, upon being asked by police officers in this way, many do not think they are free to decline. (I borrow this example from Johnson (forthcoming)).

[8] The distinction between 'goading' and 'guiding' was drawn by Stephen Darwall. (See idem (2006, 49f.)) He contends that only the latter properly grasps the nature of second-person thought: "To intelligibly hold someone responsible, we must assume that she can hold herself responsible in her own reasoning and thought." (23) For it to be possible, we need a form of 'mutual-awareness' among agents. (44). I think it expresses exactly the idea of the primary significance of self-agreement and commitment self-attribution. Related discussions cf. also Scanlon (2014, 59).

belong to the general practice of thinking and discourse, and the speaker is nevertheless subject to accountabilities and evaluations of discursive propriety. Thus, I equally see no reason for claiming that reasonable uses of imperatives cannot be brought under the framework of discursive commitment.

Conclusion

We can summarize the act-theoretic, broadly Austinian–Fregean treatment of the assertoric force I presented in this dissertation in the following two clusters of roughly related problems or notions:

(a_1) condition of demarcation →
(b_1) identification of the unity of thought through assertoric force →
(c_1) illocutionary act performing committal, assertoric content →
(d_1) first-person thought / self-agreement →
(e_1) discursive act without any known, qualifying or infelicitous conditions

(a_2) condition of distanced analyzability →
(b_2) identification of unasserted content secondary to the assertoric unity →
(c_2) locutionary act with non-committal content →
(d_2) disagreement / conceptual friction →
(e_2) discursive act with qualifying or infelicitous conditions

The intent to develop such kind of argument is to show how the two clusters of problems could finally be combined—though 'asymmetrically'—into a single account of constitutive discursive agency: On the one hand, the disagreements or the qualifications of our discursive acts are derivative of the unqualified discursive acts, i.e. acts of assertoric, performative commitment to the discursive appropriateness of the content; on the other hand, for us finite intellects, the success or the unqualifiedness of an agent's performative commitment is based on the absence of the agent's previous qualifying conditions known by the agent as the evaluative norms constitutive of her commitment. According to this view, an agent's instantiation of a norm in assertion is equally the agent's self-attribution of that norm as being constitutive of her agency—In other words, if the instantiation of a norm is part of one's discursive power, her agency can be regarded as the constitutive consciousness of the manifestation of her discursive power in bringing that norm into being authoritative.

Furthermore, I argued that our agency needs qualifications because the constitutive norms of our acts are pluralistic and socio-historically conditioned, and the success of our performances of discursive commitment is defeasible. To understand this defeasibility and thus to understand qualifications for agency, it demands a dynamic view of the constitution of discursive agency: The manifestation of an agent's discursive power is not a norm's sheer instantiation as such; rather, it internalizes the norm and thus turns the norm from being non-constitutive to being constitutive—namely, it turns the norm into the agent's know-how. Such representation of norm is a historical revisionist one, which requires both preservation and revision of the set of historical norms constitutive of one's agency—According to the dynamic view, the preservation and the revision are not two steps, neither are they two separate operations; instead, we should regard the thinker's recognitive history and its retrospective evaluation as two irreducible aspects of the same movement of disagreement resolution, belief integration, and knowledge growth, of which the thinker's discursive agency is constituted.

References

Alchourrón, Carlos and Peter Gärdenfors, David Makinson 1985. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions". In: *Journal of Symbolic Logic*, Vol. 50 (2), pp.510–530.

Almog, Joseph *et al*. (eds.) 1989. *Themes from Kaplan*. Oxford: Oxford University Press.

Alston, William 2000. *Illocutionary Acts and Sentence Meaning*. Ithaca: Cornell University Press.

Anscombe, G. E. M. 1957. *Intention*. Oxford: Basil Blackwell.

Austin, J. L. 1975. *How to Do Things with Words*. (2. Ed.) Cambridge, MA: Harvard University Press.

Bar-On, Dorit and Keith Simmons 2007. "The Use of Force against Deflationism: Assertion and Truth". In Dirk Greimann and Geo Siegwart (eds.), *Truth and Speech Acts: Studies in the Philosophy of Language*. London and New York: Routledge. pp.61–89.

Barnes, Jonathan (ed.) 1984. *Complete Works of Aristotle: The Revised Oxford Translation*. Princeton: Princeton University Press.

Beere, Jonathan 2009. *Doing and Being*. Oxford: Oxford University Press.

Bett, Richard (ed.) 2005. *Sextus Empiricus: Against the Logicians*. Cambridge: Cambridge University Press.

Borg, Emma 2004. *Minimal Semantics*. Oxford: Oxford University Press.

Boyle, Matthew 2009. "Active Belief". In: *Canadian Journal of Philosophy*, Vol. 39 (1), pp.119-147.

Boyle, Matthew 2011. "'Making up Your Mind' and the Activity of Reason". In: *Philosopher's Imprint*, Vol. 11 (17).

Boyle, Matthew 2016. "Additive Theories of Rationality". In: *European Journal of Philosophy*, Vol. 24 (3), pp.527–555.

Brandom, Robert 1994. *Making it Explicit*. Cambridge, MA: Harvard University Press.

Brandom, Robert 1998. "Action, Norms and Practical Reasoning". In: *Philosophical Perspectives*, Vol. 12 (12), pp.127–139.

Brandom, Robert 2015. *Wiedererinnerter Idealismus*. Berlin: Suhrkamp.

Brandom, Robert. *A Spirit of Trust*. Manuscript.

Brown, Jessica 2010. "Knowledge and Assertion". In: *Philosophy and Phenomenological Research*, Vol. 81 (3), pp.549–566.

Brown, Jessica and Herman Cappelen (eds.) 2011. *Assertion: New Philosophical Essays*. Oxford: Oxford University Press.

Burge, Tyler 1993. "Content Preservation". In: *Philosophical Review*, Vol. 102 (4), pp.457–488.

Cappelen, Herman 2011. "Against Assertion". In: Brown and Cappelen (eds.) 2011, pp.21–48.

Cappelen, Herman and Josh Dever 2013. *The Inessential Indexical*. Oxford: Oxford University Press.

Cappelen, Herman and Ernie Lepore 2005. *Insensitive Semantics*. Oxford: Blackwell.

Cohen, L. J. 1986. "How is conceptual innovation possible?". In: *Erkenntnis*, Vol. 25 (2), pp.221–238.

Conant, James 2004. "Varieties of Scepticism". In: Denis McManus (ed.), *Wittgenstein and Scepticism*. London and New York: Routledge, pp.97–136.

Darwall, Stephen 2006. *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.

Davidson, Donald 1966. "Truth and Meaning". Reprinted in Davidson 2001a, pp. 17–42.

Davidson, Donald 1967. "The Logical Form of Action Sentences". Reprinted in Davidson 2001b, pp.105–148.

Davidson, Donald 1982. "Paradoxes of Irrationality". Reprinted in Davidson 2004, pp.169–188.

Davidson, Donald 2001a. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Davidson, Donald 2001b. *Essays on Actions and Events*. Oxford: Clarendon Press.

Davidson, Donald 2004. *Problems of Rationality*. Oxford: Clarendon Press.

DeRose, Keith 2002. "Assertion, Knowledge, and Context". In: *Philosophical Review*, Vol. 111 (2), pp.167–203.

Dummett, Michael 1979. “What does the Appeal to Use Do for the Theory of Meaning?” In: Margalit (ed.) 1979, pp.123–135.

Dummett, Michael 1981. *Frege. Philosophy of Language.* (2. Ed.) London: Duckworth.

Dummett, Michael 1993a. “Truth and Meaning”. In Dummett 1993b, pp.147–165.

Dummett, Michael 1993b. *The Seas of Language*. Oxford: Clarendon Press.

Engstrom, Stephen 2009. *The Form of Practical Knowledge*. Cambridge, MA: Harvard University Press.

Enoch, David 2006. “Agency, Shmagency”. In: *Philosophical Review*, Vol. 115 (2), pp.169–198.

Evans, Gareth 1982. *The Varieties of Reference*. Oxford: Clarendon Press.

Evans, Gareth 1985a. “Identity and Predication”. In Evans 1985b, pp.25–48.

Evans, Gareth 1985b. *Collected Papers*. Oxford: Clarendon Press.

Fine, Kit. 2002. *The Limits of Abstraction*. Oxford: Clarendon Press.

Fine, Kit. 2007. *Semantic Relationism*. Oxford: Blackwell.

Foot, Philippa 2001. *Natural Goodness*. Oxford: Oxford University Press.

Ford, Anton 2011. “Action and Generality”. In: idem *et al.* (eds.) 2011, pp.76–104.

Ford, Anton *et al.* (eds.) 2011 *Essays on Anscombe’s Intention*. Cambridge, MA: Harvard University Press.

Franks, Paul W. 2005. *All or Nothing*. Cambridge, MA: Harvard University Press.

Frege, Gottlob 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Verlag von Louis Nebert. Reprinted in Frege 1964.

Frege, Gottlob 1880/81. “Boole’s Logical Calculus and the *Begriffsschrift*”. In Frege 1979, pp.9–46.

Frege, Gottlob 1884. *Die Grundlagen der Arithmetik*. Reprinted in Frege 1988.

Frege, Gottlob 1893. *Grundgesetze der Arithmetik*, Vol. 1. Jena: Verlag Hermann Pohle.

Frege, Gottlob 1897. “On Mr. Peano’s Conceptual Notation and My Own”. In Frege 1984, pp.234–248.

Frege, Gottlob 1906. “Introduction to Logic”. In Frege 1979, pp.184–196.

Frege, Gottlob 1915. “My Basic Logical Insights”. In Frege 1979, pp.251–252.

Frege, Gottlob 1918–19. “Der Gedanke: Eine logische Untersuchung”. In: *Beiträge zur Philosophie des Deutschen Idealismus I*, pp.58–77.

Frege, Gottlob 1964. *Begriffsschrift und andere Aufsätze*. Hildesheim: Georg Olms Verlag.

Frege, Gottlob 1969. *Nachgelassene Schriften und wissenschaftlicher Briefwechsel*, Vol. 1. Hamburg: Meiner.

Frege, Gottlob 1979. *Posthumous Writings*. Oxford: Basil Blackwell. Trans. by Peter Long *et al*. from German in Frege 1969.[1]

Frege, Gottlob 1984. *Collected Papers on Mathematics, Logic, and Philosophy*. Trans. by Max Black *et al.* Oxford: Basil Blackwell.

Frege, Gottlob 1988. *Die Grundlagen der Arithmetik*. Hamburg: Meiner.

García-Carpintero, Manuel and Josep Macià (eds.) 2006. *Two-Dimensional Semantics*. Oxford: Clarendon Press

Gaskin, Richard 2008. *The Unity of the Proposition*. Oxford: Oxford University Press.

Geach, Peter 1965. "Assertion". In: *Philosophical Review* Vol. 74 (4), pp.449–465.

Geach, Peter 1980. *Reference and Generality*. (3. Ed.) Ithaca: Cornell University Press.

Gert, Joshua 2012. *Normative Bedrock*. Oxford: Oxford University Press.

Gibbard, Allan 1990. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

Gobsch, Wolfram 2017. "Der Mensch als Widerspruch und absolutes Wissen". In: Kern and Kietzmann (eds.) 2017, pp.120–169.

Goldberg, Sanford 2015. *Assertion*. Oxford: Oxford University Press.

Grice, Paul 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Guyer, Paul and Allen Wood (trans. & eds.) 1998. *Critique of Pure Reason: The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.

Haase, Matthias 2009. "The Laws of Thought and the Power of Thinking". In: *Canadian Journal of Philosophy* 39 (S1), pp.249–297.

Haase, Matthias 2014. "Am I You?" In: *Philosophical Explanations*, Vol. 17 (3), pp. 358–371.

Hanks, Peter 2015. *Propositional Content*. Oxford: Oxford University Press.

[1] I have slightly modified the English translation of the quotes from *Posthumous Writings* in this dissertation based on the German text.

Hansson, Sven Ove 1993. “Reversing the Levi Identity”. In: *Journal of Philosophical Logic*, Vol. 22, pp.637–669.

Hansson, Sven Ove 1999. *A Textbook of Belief Dynamics*. Dordrecht: Kluwer.

Hansson, Sven Ove 2014. “Descriptor Revision”. In: *Studia Logica*, Vol. 102 (5), pp.955–980.

Hansson, Sven Ove 2016. “Iterated Descriptor Revision and the Logic of Ramsey Test Conditionals”. In: *Journal of Philosophical Logic*, Vol. 45 (4), pp.429–450.

Hansson, Sven Ove 2017. “Logic of Belief Revision”, In: *The Stanford Encyclopedia of Philosophy*. URL = <https://plato.stanford.edu/archives/win2017/entries/logic-belief-revision/>.

Hieronymi, Pamela 2009. “Two Kinds of Agency”. In: O’Brien and Soteriou (eds.) 2009, pp.138–162.

Hull, David L. 1978. “A Matter of Individuality”. In: *Philosophy of Science*, Vol. 45 (3), pp.335–360.

Johnson, Casey Rebecca (forthcoming). “Investigating Illocutionary Monism”. In: *Synthese*.

Johnson, Casey Rebecca (forthcoming). “What Norm of Assertion?”. In: *Acta Analytica*.

Kaplan, David 1978. “Dthat”. In: Peter Cole (ed.) *Syntax and Semantics*. Cambridge, MA: Academic Press, pp.221–243.

Kaplan, David 1989a. “Demonstratives”. In: Almog, Joseph *et al*. (eds.) 1989, pp.481–563.

Kaplan, David 1989b. “Afterthoughts”. In: Almog, Joseph *et al*. (eds.) 1989, pp.565–614.

Kern, Andrea 2017. *Sources of Knowledge*. Cambridge, MA: Harvard University Press.

Kern, Andrea and Christian Kietzmann (eds.) 2017. *Selbstbewusstes Leben*. Berlin: Suhrkamp.

Khoo, Justin 2017. “Backtracking Counterfactuals Revisited”. In: *Mind*, Vol. 126 (503), pp.841–910.

Kimhi, Irad (forthcoming). *Thinking and Being*. Cambridge, MA: Harvard University Press.

King, Jeffrey and Jason Stanley 2005. “Semantics, Pragmatics, and the Role of Semantic Content”. In: Szabó (ed.) 2005, pp.111–164.

King, Jeffrey and Scott Soames, Jeff Speaks 2014. *New Thinking about Propositions*. Oxford: Oxford University Press.

Kitcher, Patricia 2011. *Kant's Thinker*. Oxford: Oxford University Press.

Korsgaard, Christine 1997. "The Normativity of Instrumental Reason". In: Garrett Cullity and Berys Gaut (eds.), *Ethics and Practical Reason*. Oxford: Clarendon Press, pp.215–255.

Korsgaard, Christine 2009. *Self-Constitution*. Oxford: Oxford University Press.

Kosman, Aryeh 2013. *The Activity of Being*. Cambridge, MA: Harvard University Press.

Kripke, Saul 1977. "Speaker's Reference and Semantic Reference". In: *Midwest Studies in Philosophy*, Vol. II, pp.255–276.

Kripke, Saul 1982. *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.

Kripke, Saul 2011. "Nozick on Knowledge". In: idem *Philosophical Troubles*. Oxford: Oxford University Press, pp.162–224.

Kvanvig, Jonathan 2011. "Norms of Assertion". In: Brown and Cappelen (eds.) 2011, pp.233–250.

Lackey, Jennifer 2007. "Norms of Assertion". In: *Noûs*, Vol. 41 (4), pp.594–626.

Lavin, Douglas 2011. "Problems of Intellectualism: Raz on Reason and its Objects". In: *Jurisprudence*, Vol. 2 (2), pp.367–378.

Lavin, Douglas 2015. "Action as a Form of Temporal Unity". In: *Canadian Journal of Philosophy*, Vol. 45 (5), pp.609–629.

Leech, Jessica 2015. "Logic and the Laws of Thought". In: *Philosopher's Imprint*, Vol. 15 (12).

Lewis, David 1979. "Attitudes de Dicto and de Se". In: *Philosophical Review*, Vol. 88 (4), pp.513–543.

Lewis, David 1980. "Index, Context, and Content". In Stig Kanger and Sven Öhman (eds.), *Philosophy and Grammar*. Dordrecht: Reidel, pp.79–100.

Levi, Isaac 1991. *The Fixation of Belief and its Undoing*. Cambridge: Cambridge University Press.

Longuenesse, Béatrice 1998. *Kant and the Capacity to Judge*. Princeton: Princeton University Press.

Lowe, E. J. 2011. "How Not to Think of Powers: A Deconstruction of the 'Dispositions and Conditionals' Debate". In: *The Monist*, Vol. 94 (1), pp.19–33.

Macbeth, Danielle 2005. *Frege's Logic*. Cambridge, MA: Harvard University Press.

MacFarlane, John 2002. "Frege, Kant, and the Logic in Logicism". In: *The Philosophical Review*, Vol. 111 (1), pp.25–65.

MacFarlane, John 2003. "Future Contingents and Relative Truth". In: *Philosophical Quarterly*, Vol. 53 (212), pp.321–336.

MacFarlane, John 2005. "Making Sense of Relative Truth". In: *Proceedings of the Aristotelian Society*, Vol. 105, pp.321–339.

MacFarlane, John 2009. "Nonindexical Contextualism". In: *Synthese*, Vol. 166, pp.231–250.

MacFarlane, John 2011. "What is Assertion?" In: Brown and Cappelen (eds.) 2011, pp.79–96.

MacFarlane, John 2014. *Assessment Sensitivity*. Oxford: Oxford University Press.

Margalit, Avishai (ed.) 1979. *Meaning and Use*. Dordrecht: Reidel.

McDowell, John 1976. "Truth-Conditions, Bivalence, and Verificationism". Reprinted in McDowell 1998a, pp.3–28.

McDowell, John 1981. "Non-Cognitivism and Rule-Following". Reprinted in McDowell 1998b, pp.198–218.

McDowell, John 1982. "Falsehood and Not-Being in Plato's *Sophist*". Reprinted in McDowell 2009b, pp.3–22.

McDowell, John 1984. "Wittgenstein on Following a Rule". In: *Synthese*, Vol. 58, pp.325–363.

McDowell, John 1987. "In Defence of Modesty". Reprinted in McDowell 1998a, pp.87–107.

McDowell, John 1998a. *Meaning, Knowledge, and Reality*. Cambridge, MA: Harvard University Press.

McDowell, John 1998b. *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.

McDowell, John 1999. "Scheme–Content Dualism and Empiricism". Reprinted in McDowell 2009b, pp.115–133.

McDowell, John 2005a. "Evans's Frege". Reprinted in McDowell 2009b, pp.163–185.

McDowell, John 2005b. "Motivating Inferentialism". Reprinted in McDowell 2009b, pp.288–307.

McDowell, John 2005c. "Self-determining Subjectivity and External Constraint". Reprinted in McDowell 2009a, pp.90–107.

McDowell, John 2006. “The Disjunctive Conception of Experience as Material for a Transcendental Argument”. Reprinted in McDowell 2009b, pp.225–240.

McDowell, John 2008. “Towards a Reading of Hegel on Action in the ‘Reason’ Chapter of the *Phenomenology*”. Reprinted in McDowell 2009a, pp.166–184.

McDowell, John 2009a. *Having the World in View*. Cambridge, MA: Harvard University Press.

McDowell, John 2009b. *The Engaged Intellect*. Cambridge, MA: Harvard University Press.

McDowell, John 2010. “Brandom on Observation”. In: Weiss and Wanderer (eds.) 2010, pp.129–144.

McDowell, John 2011. “Anscombe on Bodily Self-Knowledge”. In: Ford, Anton *et al.* (eds.) 2011, pp.128–146.

Moran, Richard 2001. *Authority and Estrangement*. Princeton: Princeton University Press.

Moran, Richard 2005. “Getting Told and Being Believed”. In: *Philosopher’s Imprint*, Vol. 5 (5).

Nagel, Thomas 1997. *The Last Word*. Oxford: Oxford University Press.

Nozick, Robert 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

O’Brien, Lucy and Matthew Soteriou (eds.) 2009. *Mental Actions*. Oxford: Oxford University Press.

Parfit, Derek 1997. “Reasons and Motivation”. In: *Aristotelian Society Supplementary*, Vol. 71 (1), pp.99–130.

Peacocke, Christopher 2004. *The Realm of Reason*. Oxford: Clarendon Press.

Peramatzis, Michail 2010. “Essence and *Per Se* Predication in Aristotle’s *Metaphysics* Z.4”. In: *Oxford Studies in Ancient Philosophy*, Vol. 39, pp.121–182.

Perry, John 1977. “Frege on Demonstratives”. In: *Philosophical Review*, Vol. 86 (4), pp.474–497.

Perry, John 1979. “The Problem of the Essential Indexical”. In: *Noûs*, Vol. 13 (1), pp.3–21.

Perry, John 2009. “Directing Intentions”. In: Joseph Almog and Paolo Leonardi (eds.) *The Philosophy of David Kaplan*. Oxford: Oxford University Press, pp.187–201.

Pinkard, Terry 2002. *German Philosophy 1760–1860*. Cambridge: Cambridge University Press.

Pritchard, Duncan and Alan Millar, Adrian Haddock 2010. *The Nature and Value of Knowledge*. Oxford: Oxford University Press.

Putnam, Hilary 2000. "Rethinking Mathematical Necessity". In: Alice Crary and Rupert Read (eds.), *The New Wittgenstein*. London and New York: Routledge, pp.218–231.

Quine, W. V. O. 1951. "Two Dogmas of Empiricism". In: *Philosophical Review*, Vol. 60 (1), pp.20–43.

Quine, W. V. O. 1960. *Word and Object*. Cambridge, MA: The M.I.T. Press.

Quine, W. V. O. 1972. "The Variable". Reprinted in Quine 1976, pp.271–282.

Quine, W. V. O. 1976. *The Ways of Paradox, and Other Essays*. Revised and enlarged edition. Cambridge, MA: Harvard University Press.

Quine, W. V. O. 2000. "I, You, and It: An Epistemological Triangle". In Alex Orenstein and Petr Kotatko (eds.), *Knowledge, Language and Logic: Questions for Quine*. Dordrecht: Kluwer, pp.1–6.

Recanati, François 1987. *Meaning and Force*. Cambridge: Cambridge University Press.

Recanati, François 2004. *Literal Meaning*. Cambridge: Cambridge University Press.

Recanati, François 2007. *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press

Recanati, François 2010. *Truth-Conditional Pragmatics*. Oxford: Oxford University Press.

Recanati, François (forthcoming). "Force Cancellation". In: *Synthese*.

Rödl, Sebastian 2000. "Normativität des Geistes versus Philosophie als Erklärung". Trans. and reprinted in: Weiss and Wanderer (eds.) 2010, pp.63–80.

Rödl, Sebastian 2007. *Self-Consciousness*. Cambridge, MA: Harvard University Press.

Rödl, Sebastian 2010. "The Form of the Will". In: Tenenbaum (ed.) 2010, pp.138–160.

Rödl, Sebastian 2012. *Categories of the Temporal*. Cambridge, MA: Harvard University Press.

Rödl, Sebastian 2013. "Why Ought Implies Can". In: Mark Timmons and Sorin Baiasu (eds.) *Kant on Practical Justification: Interpretive Essays*. Oxford: Oxford University Press, pp.42–57.

Rödl, Sebastian 2014. "Intentional Transaction". In: *Philosophical Explorations*, Vol. 17 (3), 304–316.

Rödl, Sebastian 2016. "Education and Autonomy", In: *Journal of Education*, Vol. 50 (1), pp.84–97.

Russell, Bertrand 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

Scanlon, T. M. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Searle, John 1969. *Speech Acts*. Cambridge: Cambridge University Press.

Searle, John and Daniel Vanderveken 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Selby-Bigge, L. A. and P. H. Nidditch (eds.) 1978. *David Hume. A Treatise of Human Nature*. Oxford: Clarendon Press.

Shah, Nishi and David Velleman 2005. "Doxastic Deliberation". In: *Philosophical Review*, Vol. 114 (4), pp.497–534.

Shields, Christopher 1999. *Order in Multiplicity*. Oxford: Oxford University Press.

Simion, Mona 2016. "Assertion: Knowledge is Enough". In: Synthese, Vol. 193, pp.3041–3056.

Soames, Scott 2010. *What is Meaning?* Princeton: Princeton University Press.

Soames, Scott 2014. "Cognitive Propositions". In: Jeffrey King, Scott Soames, Jeff Speaks 2014, pp.91–124.

Soames, Scott 2015. *Rethinking Language, Mind, and Meaning*. Princeton: Princeton University Press.

Soames, Scott (forthcoming). "Propositions as Cognitive Acts". In: *Synthese*.

Sosa, Ernest 2015. *Judgment and Agency*. Oxford: Oxford University Press.

Stalnaker, Robert 1978. "Assertion". Reprinted in: Stalnaker 1999, pp.78–95.

Stalnaker, Robert 1999. *Context and Content*. Oxford: Oxford University Press.

Stalnaker, Robert 2004. "Assertion Revisited". Reprinted in: García-Carpintero and Macià (eds.) 2006, pp.293–309.

Stalnaker, Robert 2014. *Context*. Oxford: Oxford University Press.

Stanley, Jason 2007. *Language in Context*. Oxford: Oxford University Press.

Stanley, Jason 2011. *Know How*. Oxford: Oxford University Press.

Stekeler-Weithofer, Pirmin 2005. *Philosophie des Selbstbewusstseins*. Frankfurt am Main: Suhrkamp.

Stekeler-Weithofer, Pirmin. *Hegel's Analytic Pragmatism*. Manuscript.

Strawson, P. F. 1959. *Individuals*. London and New York: Routledge.

Strawson, P. F. 1969. "Meaning and Truth". Reprinted in Strawson (1971), pp.170–189.

Strawson, P. F. 1971. *Logico-Linguistic Papers*. London: Methuen.

Szabó, Zoltán Gendler (ed.) 2005. *Semantics Versus Pragmatics*. Oxford: Clarendon Press.

Tenenbaum, Sergio (ed.) 2010. *Desire, Practical Reason, and the Good*. Oxford: Oxford University Press.

Thompson, Michael 2008. *Life and Action*. Cambridge, MA: Harvard University Press.

Thompson, Michael 2011. "Anscombe's *Intention* and Practical Knowledge". In: Ford, Anton *et al.* (eds.) 2011, pp.198–210.

Travis, Charles 2006. *Thought's Footing*. Oxford: Oxford University Press.

U.S. v. Drayton 2002. U.S. Supreme Court.

Velleman, J. David 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Weiss, Bernhard and Jeremy Wanderer (eds.) 2010. *Reading Brandom: On Making it Explicit*. London and New York: Routledge.

Whitehead, Alfred North and Bertrand Russell 1963. *Principia Mathematica*, Vol. 1. (2. Ed.) Cambridge: Cambridge University Press.

Whiting, Daniel 2015. "Truth is the Norm for Assertion: A Reply to Littlejohn". In: *Erkenntnis*, Vol. 80 (6), pp.1245–1253.

Williamson, Timothy 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, Timothy 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

Wittgenstein, Ludwig 1921. *Tractatus logico-philosophicus*. In Wittgenstein 1984, pp.7–86.

Wittgenstein, Ludwig 1953. *Philosophische Untersuchungen*. In Wittgenstein 1984, pp.225–580.

Wittgenstein, Ludwig 1958. *The Blue and Brown Books*. Oxford: Basil Blackwell.

Wittgenstein, Ludwig 1984. *Ludwig Wittgenstein* Werkausgabe, Vol. 1. Frankfurt am Main: Suhrkamp.

Wright, Crispin 1992. *Truth and Objectivity*. Cambridge, MA: Harvard University Press.

Zagzebski, Linda Trinkaus 2012. *Epistemic Authority*. Oxford: Oxford University Press.

Zhan, Yiwen (forthcoming). "Overcoming Ossification: Hegel on Life and Rationality". In: *Hegel-Jahrbuch*.

Zhan, Yiwen. "The Efficacy of Practical Thinking". Manuscript.